DOCUMENT RESUME

ED 117 217 95 UD 015 607

TITLE	A Design to Study the Effectiveness of Individualized Instruction in the Teaching of Reading and Mathematics.
INSTITUTION	Education Turnkey Systems, Inc., Washington, D.C.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	31 Oct 75
CONTRACT	NIE-C-400-75-0073
NOTE	236p.; For "Appendices" to this document, see UD 015 608
EDRS PRICE	MF-$0.76 HC-$12.05 Plus Postage
DESCRIPTORS	*Compensatory Education Programs; Data Collection; Evaluation Criteria; Federal Programs; *Individualized Instruction; Mathematics Instruction; Measurement Techniques; National Surveys; *Program Evaluation; Reading Instruction; *Research Design; Sampling; Statistical Analysis; Teaching Methods

ABSTRACT

This report was submitted by Education Turnkey Systems, Inc. and its subcontractor, Kirschner Associates, Inc. and key University of Maryland and other consultants in response to the needs of the National Institute of Education as specified in the Request for Proposal "to develop a Design to Study Individualized Instruction, as part of a general study of compensatory education activities". The ultimate purpose of this two-phase effort is to provide both Congress and NIE with policy-relevant information on the effectiveness and effects of well-implemented individualized math and reading programs as they compare with similar standardized programs. In light of the findings which arose during the design phase, it is stated, two additional major study objects are proposed: (1) to determine the relative cost-effectiveness of the process dimensions of individualized instruction and (2) to determine the degrees to which potentially well-implemented programs are implemented during the period of observation and in turn, the relationship between degrees of implementation and program outcomes within and among programs. The proposed sampling design requires data from approximately 800 classrooms with equal or proportional representation in each of four categories, defined, for example, by whether or not specific performance objectives are assigned to each student. (Author/JM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

15607

ED117217

A DESIGN TO STUDY THE EFFECTIVENESS OF INDIVIDUALIZED INSTRUCTION IN THE TEACHING OF READING AND MATHEMATICS

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: UD
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. SP

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NIE CONTRACT No. 400-75-0073

OCTOBER 31, 1975

UD 015607

EDUCATION TURNKEY SYSTEMS, Inc.

1660 L St., N.W. Washington, D.C. 20036

ACKNOWLEDGEMENTS

In developing a design for a study as critical as this one with all its inherent technical and research problems, the practical insights, judgments, and technical skills contributed by team members were essential, if not in arriving at unanimity of opinion, then certainly in refining the critical issues. These contributions are duly acknowledged below.

Professor C. Mitchell Dayton's guidance in developing the sampling and analysis plans provided a basic structure for a design reflecting optimal trade-offs between purity and practicality. Dr. George Macready's assistance in developing outcome measures and selecting instruments contributed much to a scholarly product. Critiques of instruments by Professor William Mehrens provided useful insights and support for the proposed approach. Policy insights by Professor Gil Austin guided our conceptualization of the design. The wide and in-depth experiences of Dr. Richard Rossmiller plus his unpublished research on individualized instruction are appreciatively noted here and duly recognized in the text.

Practical insights into various aspects of our design were contributed by LEA consultants, Douglas Dalton and Sheldon Sofer. Their suggested approaches for assessing program implementation, conducting site selection, treating issues related to parental involvement, and ensuring LEA cooperation assisted in the development of a design which we feel can be implemented in the operational world of compensatory education.

Upon initiation of the design study, several individuals and groups contributed "unofficial" consultant services reflecting their interest in a highly professional evaluation of individualized instruction. Drs. Charles Willis, John Paden, and John Bahner of /I/D/E/A/ Kettering not only shared their thoughtful insights and experiences with us during the summer but provided extremely useful documentation regarding studies sponsored by /I/D/E/A/ Kettering. Dr. Marian Stearns' personal assistance, as well as her prior research endeavors, greatly assisted and influenced the proposed approach for assessing parental involvement as a separate sub-study.

The contributions of TURNKEY and Kirschner staff members are also acknowledged. Alfred Morin developed the checklist, reviewed documentation, and selected potential candidate sites. Robert Luebke provided specific critiques and assisted in the parental involvement design. July Fillmore edited the various drafts. Rosie Swart provided insight into the operational aspects of program implementation and developed the implementation assessment instruments described in the report.

And finally, we want to express our appreciation to the NIE Project Officer, Dr. Joy Frechtling, for taking the time to meet with us on numerous occassions, providing not only empathy for the problems confronting us but also insightful criticism regarding our draft design.

Charles L. Blaschke
John M. Sweeney
Blair H. Curry

TABLE OF CONTENTS

SECTION PAGE

1. INTRODUCTION

Background 1-1
Purpose of the Study 1-3
Relationship to District Survey 1 1-6
Summary of Research on Individualized Instruction 1-7

11. GENERAL APPROACH/STUDY DESIGN

Major Study Objectives 11-1
Sampling 11-2
Variables 11-2
Instrumentation 11-4
Data Collection 11-5
Data Reduction and Analysis 11-6
Underlying Philosophy of Evaluation 11-6

111. DESIGN ISSUES/TASK APPROACHES

Task 1 -- Define Program Variables 111-2
Task 2 -- Define Sampling Procedures 111-20
Task 3 -- Develop Procedures for Assessing Program Implementation 111-50
Task 4 -- Define Outcome Behaviors 111-91
Task 5 -- Develop Analysis Plan 111-126

IV. LOGISTICAL ISSUES/TASK APPROACHES

Task 6 -- Develop Data Collection Plan IV-2
Task 7 -- Provide Logistical Plan for Program Cooperation IV-20
Task 8 -- Develop Staffing Plan IV-26
Task 9 -- Develop Implementation Schedule IV-39

APPENDICES

1. Research Efforts and Findings
2. General Approach Memorandum for Parental Involvement Study
3. Proposed Design for Parental Involvement Study
4. Compensatory Education Variable Checklist
5. Sampling Questionnaires
6. Sample Size Determination
7. Selection Process for Potential Candidate Sites
8. List of Potential Candidate Sites
9. Site Identification Checklist - District Survey I Sites
10. Site Identification Checklist - Non District Survey 1 Sites
11. Program Implementation Instrument Design
12. Draft Implementation Assessment Instrument
13. Suggested Interview Instrument
14. Data Collector's Manual

SECTION I

INTRODUCTION

Education Turnkey Systems, Inc. and its subcontractor, Kirschner Associates, Inc. (KAI) and key University of Maryland and other consultants are pleased to submit this report in response to the needs of the National Institute of Education as specified in the Request for Proposal (NIE-R-75-0022): To develop a Design to Study Individualized Instruction, as part of a general study of compensatory education activities. The ultimate purpose of this two-phase effort is to provide both Congress and NIE with policy-relevant information on the effectiveness and effects of well-implemented individualized math and reading programs as they compare with similar standardized programs.

BACKGROUND

The Education Amendments of 1974 provide for a thorough study of compensatory education.* Specifically, Clauses (1) through (6) of Section 821 (a) of P.L. 93-380 mandate the National Institute of Education to conduct a study which will include: (1) an examination of the purposes and the effectiveness of such programs; (2) an analysis of the means to identify the children in need of such programs ; (3) an analysis of the effectiveness of methods and procedures; (4) an examination of alternative methods; (5) the administration, by NIE, of not more than 20 experimental programs; and (6) findings and recommendations.

The Research Plan, Compensatory Education Study (16 December 1974, pp. 3-4) describes the three basic components of the overall study (of which District Survey II as modified is a part) as:

---

* See TURNKEY Proposal, June 6, 1975, hereafter referred to as Proposal.

> "Research on Programs for students in need of compensatory education....
>
> "Research on the allocation of funds to support compensatory education programs...
>
> "Research on the operation of federal, state, and local agencies which deliver compensatory education programs."

The large-scale NIE study of compensatory education is unique in several important respects --- specificity, tone, and context. Congress has mandated the overall study in unusual specificity: a) policy research directed to specific questions or issues; b) prior Congressional review of the proposed design and plan; c) solicitation of advice from the National Advisory Council on Education of Disadvantaged Children in the design and operational phases; and d) restrictions regarding review of study findings prior to submission to Congress. Hence, the mechanics and procedures of this study differ significantly from others conducted in this field.

Implicit in these specified characteristics is a unique tone --- the desire for a comprehensive study of compensatory education more relevant to Congressional needs. During the hearings and subcommittee sessions on the 1974 Amendments to ESEA, several different approaches and techniques for compensatory education received much more interest than in the past, including: a) educational deficiency as measured by criterion-referenced tests as the basis for allocation of funds; b) contingency or incentive funding based upon program or student achievement; c) alternative procedures and mechanisms for administering/monitoring compensatory education programs; and d) evolution of national standards of educational excellence and equity beyond the concept of equal opportunity. Policy and disciplinary research findings emanating from this study should provide useful information for the policy formulation process.

The unusual tone of this study can be attributed to several reasons, including frustrations of key Congressmen due to the lack of a consensus of which programs work best and why, and confusion (and hence, criticism) as to who should demonstrate results --- local education agency (LEA) program staff or those conducting evaluations? An awesome responsibility is presented to NIE and its contractors: this particular research effort must be as scientifically flawless as it is feasible in design and execution, so that any lack of evidence of success can be attributed to programs or their implementation, not to the research effort itself.

The context of the overall study differs from most others. Supported by the results of recent state education agency compensatory education studies in California, Michigan, and elsewhere, and other information which suggests that compensatory education can work if it is individualized, Congress has clearly expressed its intent for the development and implementation of individualized programs of instruction for students. A study of this assumption must, therefore, be carefully designed, comprehensive in scope, cautious in attribution, and insightful in suggesting feasible ways for further improvement.

PURPOSE OF THE STUDY

In the RFP, NIE has specified the overall purpose and characteristics of this study:

> "...to compare for compensatory education programs teaching reading and mathematics, the effectiveness of well-implemented individualized instructional programs with that of well-implemented standardized programs. The focus of the study will be on assessing the effects of individualization and not on the variety of additional dimensions along which programs may vary. It will employ

> in-depth observations of programs to fully describe their operating characteristics, to determine the degree of implementation, and to describe settings where adequate implementation may be difficult to achieve. Programs will be assessed by measures of reading and mathematics achievement and by their broader effects on classroom environment."

Individualization of instruction has been identified for specific attention because of the expansion of the concept in education generally, enthusiastic support of education researchers and practitioners, and interest shown by Congress. Clause B of Section 821 (a) of P.L. 93-380 mandates NIE to conduct "an analysis of the effectiveness of methods and procedures for meeting educational needs of children, including the use of individual written educational plans for children." The House Committee on Education and Labor Report to Accompany HR 69 (21 February 1974) stated the intent of Congress regarding individualized instruction very clearly:

> "The Committee Bill amends Title I to include a sense of Congress statement that local school districts are encouraged to develop individually described plans for all children participating in Title I programs. These plans are not meant to be required under Title I but the adoption of this statement is meant to show the feeling of Congress that there is a need for more individualized education and a need for more systematic planning in education to include the involvement of teachers, parents, and child."

It should also be noted that in the minority dissenting views, no member of the Committee took exception to the "sense of Congress" as stated above, even though other issues, such as sex education, textbook censorship, parental involvement, and other sections of the 1974 Amendment, were debated. In light of this apparent unanimity it is obvious that Congress has a commitment to individualized instruction and wants to learn more about its effectiveness and the conditions under which it helps disadvantaged students most.

Some of the more important policy questions to be addressed which emerge upon a review of the RFP and the Hearings and Testimony leading to the passage of the 1974 Amendments are as follows:

- How effective are well-implemented individualized programs in raising achievement in math and reading for disadvantaged children? The unit of comparison would be well-implemented standardized or traditional type compensatory education programs with similar objectives and similar compensatory education students.

- What is the impact and interrelationship among other effects created in the classroom and school environment that can be attributed to the effective implementation of individualized instruction? These general effects would include teacher morale and expectations, sense of involvement, classroom discipline, student attitudes, and time spent on productive activities, among others.

- To what extent do local education agencies utilizing well-implemented individualized programs of instruction meet the specific needs of individual students? What types of encouragement or what types of barriers need to be removed in order for LEAs to do this job more effectively? As the House Committee Report noted, "The Committee feels strongly that the local school agency is the appropriate level to determine the specific needs of educationally deprived children and should be primarily responsible for determining approaches to meet those needs" (page 21).

- To what extent does individualized instruction facilitate or provide an atmosphere conducive to the furtherence of additional Congressional intents such as: a) the degree, extent, and types of parental involvement; b) the targeting of compensatory education funds on schools with high concentrations of eligible students; c) the facilitation of higher quality evaluations to provide feedback for programs improvement purposes; and d) the facilitation of more effective and higher quality planning and directly related teacher training at the LEA and building level?

Perhaps as important as the issues to be addressed in this study are those which are not to be addressed. For example, this study will not attempt to determine the degree to which individualization exists nationally in compensatory education programs nor the type and nature of these programs. Nor is this study designed to analyze and rank the various individualized instructional "packages" and curriculae in terms of their relative cost-effectiveness.

Since this study is part of the overall study of compensatory education conducted by NIE and is related to District Survey I, a discussion of the relationship between the two will assist in focusing upon the specific issues to be addressed.

RELATIONSHIP TO DISTRICT SURVEY I

District Survey I is designed to be a general survey of compensatory education activities in a sample of districts, examining: a) the purposes of compensatory education programs, as stated and perceived at all levels; b) the planning and coordination mechanisms used to address these purposes; c) operating characteristics and procedures followed in meeting objectives; and d) evaluation activities and techniques used. The major questions to be researched, analyzed, and answered in District Survey I are:

1. What are the purposes of existing compensatory education (comp ed) programs, as stated and perceived by education agencies, staff, and parents?
2. What types of planning procedures and coordinating mechanisms are used to ensure efficient use of funds and program integration into regular school activities? To what extent have LEAs developed programs consistent with and relevant to stated goals and objectives? Are objectives realistic and programs feasible?
3. What are the levels of funding and the nature of the allocation of these funds? What are the operating characteristics of the various types of comp ed programs as they reflect instructional coordination mechanisms, training, nature and extent of staff participation, and associated problems?
4. What is the nature and extent of evaluation activities prescribed and used to assess program success, including type (i.e., summative or process), criteria and measure, design (e.g., pre/post, post-test only), type of evaluation (e.g., internal, third party), feedback (e.g., extent and use), and reporting (e.g., utility, timeliness, relevancy)? How successful have various types of programs been in making modifications in response to LEA and SEA evaluation results?

The purpose of this study is to assess the effectiveness of individualized instruction as it is currently used in compensatory programs in schools. The study is not meant to be a comprehensive survey like District Survey I, but rather a concentrated examination of selected programs which provide inidvidualized classroom instruction. The major focus will be on assessing program success in the areas of reading and mathematics achievement, and general implications of individualized instruction for the classroom environment, while District Survey I has a much broader scope.

However, as explained later, NIE has requested that the design of this study be coordinated with that of District Survey I regarding: a) coordinated site selection; and b) "validation" of data collection instruments. In order to ensure effective coordination, as described in our Proposal, differences in definitions of compensatory education "programs" and "student" (p. I-8), "educational plans" (p. I-7), and "individualized plans of instruction" (p. I-9) will have to be resolved. Moreover, the overall time schedules of the two studies must be compatible. These considerations are discussed in greater detail in the task descriptions in Section 3.

SUMMARY OF RESEARCH ON INDIVIDUALIZED INSTRUCTION

A. Status of Individualization of Instruction

In order to assist in refining the overall design a rather extensive review was conducted to determine the status of research on individualized instruction. While this survey and research was not as intensive as was desirable due to the relatively short timeline, more than 75 reports and documents were reviewed and discussions were conducted with over 50 individuals who have been intimately involved in the development and implementation of individualized instructional programs across the country. Below, we summarize the results and findings of this survey, identifying the implications for the overall study design and the design approach we are proposing in the subsequent chapter. In those instances where the findings noted below were stated in greater detail in our original Proposal, we have so indicated with appropriate references.

1. While the philosophical differences between individualized and standardized instruction are generally easy to describe, the philosophical differences among individualized programs in math and reading are significant.

Traditional or standardized instruction perceives the class or group as an entity itself. Each student is presumed to have relatively equal learning needs, abilities, and responses. Instruction is teacher-paced and scheduled to meet the convenience of the school and the teacher. Students taught by traditional or standardized techniques are generally given the same assignment regardless of individual capabilities or progress. Where individual assignments are made they are usually to be completed in a specific time period.

On the other hand, the philosophy underlying individualized instruction recognizes that each individual learns his/her own way, and his/her own pace. Individualized instruction therefore seeks to motivate the child by finding areas of interest and prescribing individual assignments based upon the child's interest areas. As a student-oriented approach, it requires a diagnosis of individual students to determine the kind of learning experiences needed and then to prescribe programs that are mostly pupil-directed and pupil-administered within general limitations.

In reviewing the various instructional systems and critiques of many of the same systems conducted by others (e.g., DeVault and EPIE), the philosophical foundations underlying the various individualized learning systems within the context of the distinction made above do vary, if not in kind then certainly in degree.

First, the locus of control and decision-making underlying several individualized learning systems varies considerably. Several Follow Through models (e.g., Far West Regional Education Laboratory), several contingency-based models (e.g., Grand Rapids Contract Learning Project), and, to some extent, "packaged" systems (e.g., High Intensity Learning System), allow for greater student involvement in: a) selecting instructional objectives and/or activities including; b) selecting reinforcement activities; c) accepting responsibility for pacing; and d) recording progress. Other systems (e.g., the University of Oregon Follow Through model, New Century), which require intensive pre-service training, are designed to insure greater control by the teacher in activities such as: a) prescribing specific activities and enrichment materials; b) instructional management, including record keeping; and c) applying motivational techniques.

Second, approaches to individualing classroom instruction also vary. The IGE process represents an illustration of the "top-down" approach for individualizing math and reading programs through an extensive preparation phase which includes participation in a national "network" or "league" creating change in the district and school environment. The specific individualized instruction program, or the one which is developed by the staff, is of lesser importance than creating an environment conducive to individualization. On the other hand, learning systems such as BRL and New Century utilize a pilot program approach in individual buildings with the focus on individualizing instruction in specific classrooms ("bottom-up" approach) and over time changing other environmental conditions and factors which can facilitate a more effective implementation of the specific program. While over time many of the same problems are confronted and eventually solved, the immediacy of resolution varies.

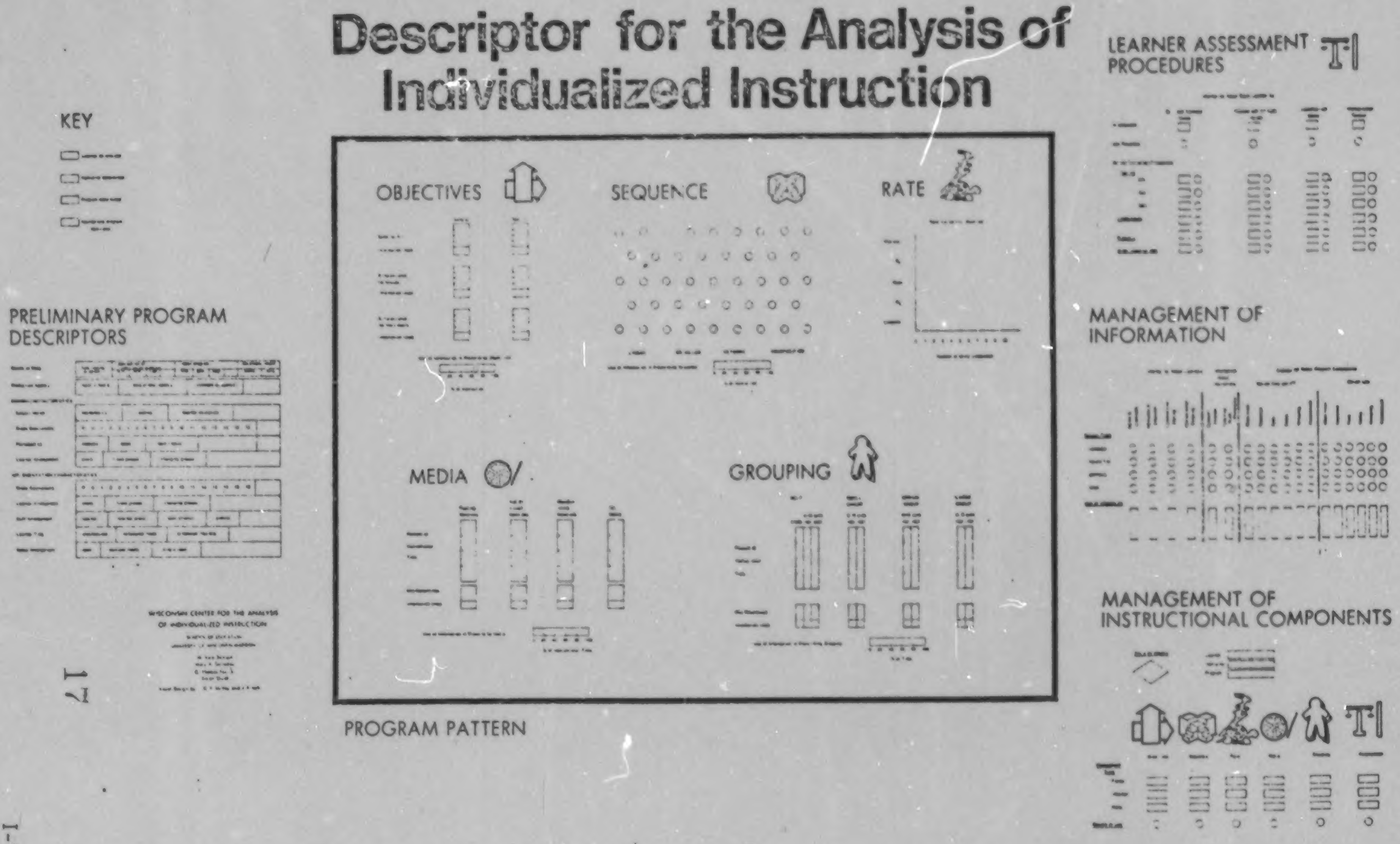
2. Definitions of individualized instructional programs vary in terms of scope and specificity; existing classification schemes also vary accordin to intended purpose.

During early review of the literature on individualized instruction and subsequent discussions with LEA staff and technical consultants, it became apparent that no uniform definition of individualized instruction is accepted. Moreover, any definition to be used in the study would have to be used for the purposes of the study, identifying those characteristics which would be the focus of the study. The RFP proposed several characteristics of individualized programs, including the existence of stated performance objectives, special procedures for diagnosis and prescription, the use of progress tests, and, most critically, student progress in remedial instruction which is based on student performance on progress tests and allows students to proceed at their own rate.

One of the first attempts to develop a set of procedures and classification scheme was developed by DeVault, et.al. (1973); a chart describing this descriptor is included in Exhibit I-1. After field testing, he indicated that the instrument was useful in describing programs as they exist and the degree to which individualization in a program could be described in accordance with standards established by the individual teacher. The instrument is limited, however, in comparing various programs in terms of individualization. Since the early 1970's, EPIE has developed two instruments, one a hybrid version of the other, to classify programs along 16 items, providing a range of variables (EPIE, 1972, see Exhibit I-2). Subsequently, a variation of this classification model was developed and applied to over 25 learning systems in accordance with the characteristics described in the implementation guides (1975). The former instrument was designed to be used by practitioners in assessing the degree to which various types of materials could be used for individualized instruction, while the primary purpose of the latter was to describe the characteristics of over 25 learning systems.

A review of these classifications/descriptors, as well as others (e.g.,/I/D/E/A/'s IGE Implementation kit), indicates that while it is possible to develop specific descriptors and procedures for classifying program characteristics, such descriptors must focus upon the specific purpose in mind: hence, while the above documents are extremely useful, a classification scheme specifically designed for the purposes of this study was needed. This scheme

Exhibit I-1



Descriptor for the Analysis of Individualized Instruction

KEY
PRELIMINARY PROGRAM DESCRIPTORS
OBJECTIVES
SEQUENCE
RATE
MEDIA
GROUPING
PROGRAM PATTERN
LEARNER ASSESSMENT PROCEDURES
MANAGEMENT OF INFORMATION
MANAGEMENT OF INSTRUCTIONAL COMPONENTS
WISCONSIN CENTER FOR THE ANALYSIS OF INDIVIDUALIZED INSTRUCTION

Exhibit I-2

Classification and Range of Variables for Assessing Individualized Instructional Materials

Constructs and Variables	Product Characteristics		
	X	XX	O
A. OBJECTIVES			
1. Selection of Objectives	Explicit	Vague	Absent
2. Specification of Outcome Goals	Fixed	Variable	
3. Population Specific	Fixed	Variable	
B. ORGANIZATION			
4. Structure of Subject Matter	Defined	Implicit, Not Discernible	
5. Sequence	Fixed Entry	Differentiated Entry	
6. Scope			
6.1 Coverage	Fixed Coverage	Flexible Coverage	
6.2 Range	Narrow Range	Broad Range	
7. Branching	Linear Route	Multiple Route	
8. Recycling	Limited	Multiple	
9. Selection of Materials			
9.1 Supplementary Materials	Limited	Many	
9.2 Selection of Materials	Explicit	Open	
10. Learning Environment	Explicit	Implicit	
C. METHODOLOGY			
10. Learning Environment	Closed	Open	
11. Methodology	Structured	Eclectic	
12. Time	Fixed	Flexible	
13. Pacing	Fixed	Variable	
14. Modes	Limited	Multiple	
D. EVALUATION			
15. Evaluation of Learning	Norm-Referenced	Criterion-Referenced	
16. Evaluation Approaches	Teacher-Centered	Pupil-Centered	Absent

EPIE Report, Number 46

is described in Section III/Tasks 1 and 2 of this Report.

3. Operating characteristics vary considerably among individualized learning systems, either by design or for other reasons.

While this finding is implicit in our earlier observations, the nature and degree of variation in critical areas is apparent from a review of implementation guides and observations of a number of operating programs. And as noted below, these differences have major implications for the study design. The primary implication for this study is the need to "capture" the qualitative differences among programs, especially in terms of implementation variables. Five principal categories encompass these operating characteristics:

a. Diagnosis/Prescription

One critical element in the individualization process is that of diagnosis/prescription. In practice, procedures vary. "Tests" that are administered may be subjective tests, standardized norm-referenced, or criteria-referenced tests, specifically designed for the program. Analysis and interpretation can be based upon test scores or subjective evaluation. Prescription of instructional approaches and materials may be based upon a pre-programmed procedure (e.g., University of Kansas Follow Through, Project PLAN) which can be computer-based or can be developed more subjectively by individual teachers or groups of teachers (e.g., Grand Rapids Project Target). In most instances some subjective elements based on teacher interpretations are followed.

b. Mastery/Proficiency Levels

In the vast majority of individualized programs, mastery or proficiency levels are prescribed, recommended, and/or developed in order to determine how well students have mastered particular materials prior to branching or continuing the prescribed course of instruction. Mastery tests or exercises may be either "paper and pencil" or "performance-based," covering specific learning objectives or skills such as word attack, decoding, etc. Specified criteria for mastery may include a specific number of items for a skill or performance objective (e.g., three out of four items). At the lower grade levels (i.e., K-1), the number of items per objective will usually be fewer than at the higher elementary grade levels. In other instances, the mastery level may require satisfactory completion of all items associated with that skill or learning objective. In some programs, up to three forms of a particular set of items will be developed for each objective, for diagnosis, progress checks, and "post-test" reporting. Many factors influence the mastery or proficiency levels used. Where minimal performance objectives are specified by the SEA or district and contingencies such as subsequent funding are based upon the "achievement of student" (e.g., in the Michigan Chapter 3 Compensatory Model), the number of items to be successfully completed in determining mastery will be relatively few. Conversely, when funding contingencies do not exist and teacher aspirations are higher, such as in the initial year of implementing a teacher-developed individualized program, the number of objectives to be mastered as well as the criterion levels will often be unrealistically high. During the program selection and analysis phase of this study, these variances will be addressed.

c. Degree of Teacher Discretion

The degree to which the program is designed to be "teacher proof" varies among programs. In certain types of individualized programs (e.g., New Century), the implementation plan is extremely specific as regards the utilization of materials and teacher self-evaluation instruments. In others (e.g., Project PLAN), teacher training is limited to the implementation of its specific programs. At the other end of the continuum are programs (e.g., Grand Rapids Project Target) which merely structure the nature of teacher involvement in designing and implementing the instructional program. For example, the "system" might consist of an extensive teacher training program and a taxonomy of objectives keyed to a variety of source materials (e.g., HILS). The teacher is trained to design the program, select and categorize materials, develop and/or modify tests, etc. The variances described above raise critical issues in determining the degree to which an individualized instructional program is well-implemented. For example, in the New Century program, a baseline for determining deviations from the suggested implementation plan would be easy to identify. However, in other programs with very general implementation procedures leaving much discretion to the teacher, there exists very little baseline in determining the degree to which the program is actually implemented in accordance with a plan. This issue is addressed in our sample design discussion.

d. Content Sequences

Evaluators attempting to compare the effectiveness of various individualized programs have often found that different programs emphasize the development of different skills at different levels with different degrees of intensity. In mathematics, the Gatteno "adaptive approach" is based on individual learning styles, directed by a highly trained teacher. Over a short period of time it would be difficult if not impossible to compare this program with the BRL math program, for instance, which is structured and sequenced differently. In reading, certain systems emphasize the development of specific skills at specific grade levels over a very short but intensive period whereas others develop a number of skills almost concurrently over a longer period of time. Since the content and sequencing of instruction vary among programs, care must be taken to ensure that these considerations are taken into account in the selection of sample and instruments.

e. Classroom Management/Organization

Classroom management and organization of individualized programs will vary from very structured, inflexible center designs to unstructured almost indescribable situations such as those in open classrooms or open schools. The results of the PREP study conducted by the USOE in the late sixties found general methods and organizational patterns in the 46 schools surveyed, including: directed activities in multiple or single learning areas under a regular or unscheduled time period; and selected activities in multiple or single areas with time scheduled or unscheduled. Classroom management techniques which have more recently emerged include: computer-managed instructional programs where diagnosis and prescription are conducted outside the district with immediate feedback to the teacher and often the student (e.g., Capital Areas Skill Center reading program, Project SCORE); and student-managed programs where students keep their own records of mastery and the teacher merely spotchecks. In most cases the nature of the classroom management/organization of the individualized programs is dependent upon or influenced by: the suggested design inherent in the individualized program itself; local facilities and building considerations; and the nature of the instruction (e.g., high intensity remedial vs. long term developmental). In many instances the organizational and management characteristics reflect the desire to avoid barriers which otherwise might impede the effective implementation of individualized instruction.

4. The majority of individualized "systems" have undergone or are presently undergoing change.

The information gathered by project staff during the design phase of the study indicates that most individualized instructional systems have undergone or are presently undergoing significant changes in content, structure, and in certain instances, philosophy.

The nature and reasons for these changes vary among the categories of projects.

Publishers appear to be attempting to simplify the implementation of the various systems. They are also recognizing that teachers, where economically feasible, desire a variety of educational materials. Hence, several (e.g., HILS) have developed taxonomies of skills keyed to materials other than the ones that they specifically publish. Moreover, what might have been advertised as a complete and total learning system several years ago has now been broken up into different "products". For example, Westinghouse Learning Corporation will now provide the performance objectives, the test program (SCORE), as well as the teaching-learning units, as individual "products" emanating from Project PLAN. HILS includes not only a taxonomy of objectives and source materials but also provides services whereby a district can purchase a computer printout of the specific objectives covered by the materials in a specific classroom. Publishers such as Prescriptive Learning, Inc., Education Projects Corporation, and Zweig Associates (Fountain Valley) provide similar differentiated products.

Instructional programs have also become less programmed in nature, allowing for more teacher discretion during implementation. The reasons for these trends reflect a combination of increased sophistication on the part of the client (e.g., more reading specialists involved in decision-making), general economic constraints facing schools, and greater teacher involvement in decision-making generally at the building level.

A review of materials and discussions with Follow Through Model directors and responsible USOE officials indicate that virtually all Follow Through Programs have recently undergone or are presently undergoing changes. Indeed, one of the difficulties in designing the SRI Follow Through evaluation was an attempt to determine baseline regarding the implementation procedures. The reasons for these changes include: relevant program feedback, unexpected availability of funds provided to LEAs, and lack of sponsor technical assistance and training (e.g., several Follow Through developers have gone out of business). The only exception to the rule appears to be the LRDC Follow Through model which according to the Project Director, has not undergone any significant change in the past two years.

The most stable group of individualized projects appear to be those used in Federally and state validated models which have been submitted (or are presently being submitted) to the Dissemination Review Panel at USOE. The changes which have occurred in these programs have been based upon program feedback and to some extent upon general economic conditions and districtwide policies. Stability at sites nominated by publishers as being well-implemented appears to be basically a function of the stability of districtwide policies and procedures, particularly in districts which are using many different programs (e.g., Dallas, Dade County). In these districts the most disruptive elements effecting program stability appear to be implementation of court-ordered desegregation plans through bussing or transfer of staff.

The implications of these factors on sample selection and initial verification for program stability appears to be significant indeed, as described in the following Section.

B. Effects and Effectiveness of Individualized Instruction: Preliminary Findings

Below, we briefly review some of the most relevant research on individualized instruction, focusing upon: a) the relationship between school and classroom environment and individualized instruction, including staff, student, and other attitudes; b) the effectiveness of individualized programs in math and reading; and c) the relationship between program cost and effectiveness. In our Proposal, we summarized a number of large-scale studies and research efforts and findings related to compensatory education in general, some of which are relevant to subsequent discussions in our design. (see Appendix I.)

1. Most research on the effects and the effectiveness of individualized instruction have been conducted by groups or individuals involved in the development or implementation of the programs.

With the exception of a number of doctoral dissertations and independent evaluations funded by groups such as /I/D/E/A/ Kettering, the majority of research on individualized programs has been conducted by people involved in the development or implementation of a program. Although the standards of the research have generally been of high professional quality, very few (if any) comparative studies of individualized programs (with the possible exception of SRI Follow Through Study) have been made. Hence, many of the findings below have limited generalizeability regarding "individualized instruction".

2. Signficant school environment or climatic changes appear to be associated with the implementation of individualized instructional programs.

The relationship between school climate or environment and individualized instruction has been the research focus of a large number of studies sponsored by /I/D/E/A/ Kettering; however, such evaluations have been limited to schools participating in the /I/D/E/A/ "IGE Network". In addition to evaluations of the IGE program, the Wisconsin Research and Development Center for Cognitive Learning has evaluated and otherwise sponsored research on the Wisconsin Design Reading Program and other curriculum developed by the Center (e.g., DNP and IGM), as well as IGE. Other studies tangentially focusing on this issue include the SRI Follow Through Study, and studies sponsored by Westinghouse Learning Corporation (regarding Project PLAN) by New Center.

Summarizing the impact of individualized instruction, particularly the Wisconsin design, Rossmiller ("Research and Evaluation Related to IGE", Draft, July 1975) summarized the findings of Gresso which are representative of other studies of IGE:

> "High-implementation IGE schools were significantly more open-climated and familiar-climated than low implementation schools. Low implementation IGE schools were significantly more closed-climated than high implementation schools."

IGE schools are characterized by: the schools having a greater degree of openness, the schools having more open autonomous climate, teachers having a greater degree of esprit de corps, principals having greater trust; principals having greater consideration...in contrast to high-implementation IGE schools, low-implementation schools are characterized by: the schools having a more paternalistic, closed climate; teachers having a high degree of hindrance; principal having a high degree of aloofness.

Other relevant findings include: a) multi-unit IGE schools are more significantly achievement, change, and science oriented; b) teachers perceive school climate being more open in IGE schools vs. non-IGE schools (Kelly, et al.); c) that self-concept of students in IGE schools is significantly higher than the self-concept of students in non-IGE schools (Nelson); d) that the environment of IGE schools was conducive to the development of favorable student attitudes towards peers (Edwards). As Rossmiller concludes:

> "In general, research has shown the climate of multi-unit schools to be significantly different from the climate of non-multi-unit schools. In terms of organizational climate, multi-unit schools are more open; are more change and achievement oriented; possess staffs that are instructionally satisfied, more motivated, and perceive greater levels of productivity; and employ more educationally progressive practices. In terms of learning climate, multi-unit school students have more positive self-concepts, more favorable attitudes towards fellow pupils, and more favorable attitudes towards school."

In terms of organizational patterns and decision-making at the building level, findings from numerous studies including the TURNKEY Study of Compensatory Education in Michigan, indicate that individualized instructional programs are associated with: a) decentralized decision-making through a participatory process (TURNKEY, Walter, Smith, Gramenz); b) the role of the principal as a facilitator, monitor, and leader (TURNKEY, Richardson); c) a flexible and less rigid organizational structure allowing for shared decision-making and coordination (Pelligren, Herrick, and TURNKEY).

While the variables discussed above seem to be significantly related to the introduction and/or implementation of individualized instructional programs, other studies have made additional findings of interest to this study. Kelley, in his study of the relationship between individualized instruction and school climate, found that a significantly stronger relationship between climate changes and the introduction of individualized instruction occurs in rural and inner-city LEAs. In addition, he found that climate changes increased from the first to the second year; but in the third year tended to regress to a rigid, less open climate. While open climates and flexible organizational structures characterize well-implemented individualized schools, Walter found that there existed stricter adherence to instructional accountability procedures in multi-unit schools vs. traditional ones. In addition, while greater role differentiation exists, school staff indicated a perception of closer personal relations among staff.

The results of the TURNKEY Michigan study indicates that a significant relationship between individualized instruction and school climate exists when compared with the school climate of traditional or non-individualized programs. Many of these variables are outlined in the discussion of program variables in Section III of this report.

3. Existing research indicates that the introduction of individualized instruction has a significant positive impact upon student growth in math and reading (although the units of comparisons can, in many instances, be questioned.)

In 1973, USOE's Division of Compensatory Education conducted a synthesis of six major studies related to compensatory education:

- Strategies of Compensation: A Review of Educational Projects for the Disadvantaged in the United States (Organization for Economic Cooperation and Development, 1971).
- Compensatory Education: Evaluation in Perspective (Edmund W. Gordon, Information Retrieval Center on the Disadvantaged, 1970).
- How Effective Is Schooling? A Critical Review and Synthesis of Research Findings (Final Report to the President's Commission on School Finance, Rand Corporation, 1971).
- ESEA Title I: A Reanalysis and Synthesis of Evaulation Data from Fiscal Year 1965 through 1970 (American Institutes for Research, 1972).
- Draft, Final Report, Exemplary Projects Studies (Columbia University, 1972).
- State Title I Evaluation Reports for FY 1972.

In an attempt to identify common characteristics of effective programs, USOE identified several characteristics of individualized programs of instruction. As summarized in a NSPRA Report, these characteristics included:

- "Clear objectives which must be clearly written and stated in specific measurable terms; instructional techniques and materials must closely relate to those objectives."
- "Attention to individual needs which includes a careful diagnosis and individual plan for each student."
- "Flexibility and grouping which allow staff opportunities to provide small group instruction and to teach frequently on a one-to-one basis. USOE notes that when group instruction was part of the daily program it tended to be more effective if students were not confined to the same group for more than several days without reassessment of the teacher's and students' strengths."
- "Personnel management which allows key staff personnel to work individually with teachers in the classroom. USOE stresses the need for much coordination and cooperation among staff and a well designed inservice program."
- "Structured program approach which stresses sequential order and activity. Pupils must also receive frequent and immediate feedback."

In 1973, Mayfield conducted a comparative study of conventional and individualized mathematics programs, using the MAT to assess the growth of 40 matched pairs of students. While the student gains were not statistically significant overall, the individualized programs demonstrated slightly higher gains in problem solving and computation. Significant gains in self-concept, measured by the Piers Self-Concept Scale, were found. Lazich's evaluation of the Wisconsin design for grades K-3, found statistically significant differences at the third-grade level in word attack skills when compared to a control group. A summary of the effectiveness of the Wisconsin design and other instructional procedures and curricula developed by the Wisconsin Research and Development Center synthesized the findings of more than fifteen studies. In all instances, student progress in cognitive areas, as measured by "embedded" criterion-referenced tests built into the program, indicated student growth particularly in the area of word attack skills. In addition, comparative studies between the Wisconsin design and the Fountain Valley Teacher Support System and Prescriptive Reading Inventory indicated that students enrolled in the Wisconsin design program achieved significantly higher mean ratings, again measured by criterion tests.

In Michigan, over the last three years, rather extensive evaluations have been conducted of individualized programs implemented under the rubric of "performance contracting". Based on norm-referenced standardized tests, student performance in cognitive areas in Menominee and Sault Ste. Marie projects, involving five to ten buildings each, averaged 1.5 and 1.9 months gain per month in the program over the last three years. Menominee has developed an individualized mathematics program, while in Sault Ste. Marie, the reading program has been individualized. Only in the Menonimee site was a control group used for comparison. This comparison of gains showed only slightly higher gains in the experimental group (Bryan, 1974). A review of the 50 projects approved by the Dissemination Review Panel (DRP) which utilizes "educational" gains as a major criterion, found that approximately one-third of the projects were individualized, with the remainder ranging from only slightly individualized to relatively standard programs. In most instances the unit of comparison was expected growth, measured by norm-referenced standardized tests or normative comparisons by grade level. While control groups were used for comparisons in a limited number of the evaluations, in no instance did the documentation reveal that random assignments to treatment and control groups had been used. Indeed, as DRP projects were assessed by RMC as potential PIP sites, many DRP sites were dropped due to various external "threats" to the evaluation designs. (Tallmadge, 1975).

4. Attempts to relate cost to the effectiveness in assessing individualized instruction and/or comparing it with "traditional instruction" have been minimal and simplistic in design.

The vast majority of the "cost-effectiveness" studies of individualized education have been conducted or sponsored by model developers and publishers, or individual districts. The findings are usually reported in a simplistic cost-per-unit-gain, determining the cost associated with specific variables which contribute to performance. Typical studies include: a) positive cost-effectiveness results in the Cincinnati Public Schools through the use of the New Century program; b) a comparative cost-effectiveness analysis of three types of individualized program in Dade County conducted by its Division of Research and Evaluation; c) a similar cost-effectiveness study of five types of reading programs in the Dallas Independent School System conducted by its Division of Research and Evaluation; and numerous studies by publishers such as Random House, Betti-Kit, McGraw-Hill, SRA, and others. Aside from their simplicity, for the most part these studies fail to capture all of the relevant resources, but rather focus on out-of-pocket or operating costs related to the specific program.

In 1973 Boardman and Hudson developed a cost-analysis model to identify cost factors directly associated with the implementation of IGE. The instrument for data collection consisted of self-reporting, personal interviews, and on-site visitations. Similar to many of the evaluations conducted by publishers, this model attempted to identify the cost associated with implementing IGE as a process (as opposed to the instructional programs used in the IGE process); however, it did not attempt to relate total cost or the cost of specific activities to any outcome measures.

Several attempts over the last few years have been made to develop models to assess the cost and effectiveness of compensatory education programs. Since about 1971, the Education Testing Service has been conducting a study of compensatory education reading programs to identify possible effects on the development of reading skills in the elementary grades. One phase of the study includes a comprehensive evaluation of reading program characteristics and attempts to analyze the cost-effectiveness of certain variables. Based on discussions with individuals involved in this study, several relevant issues are worth noting:

- The definition of a program (i.e., similar studies receiving similar treatment) has created problems during the project's analysis phase because of the wide variety of "programs".
- Success criteria and test administration procedures are being questioned in light of the occurrence of "bottoming" and "topping" affects, creating difficulty in assessing actual vs. expected gains. The model used is similar to the RMC Model developed several years ago and uses standard pricing. It does not include a "trade-off" capability which is highly desirable for assessing individualized instruction. A similar model is now being designed for use in the cost-analyses of Follow Through and the "Sustaining Effects" Study (USOE).

Since 1973, TURNKEY has been conducting a cost-effectiveness study for compensatory education programs for the state of Michigan. Although the program did not focus solely on individualized instructional systems, the results provide useful insights into the variables associated with effective programs and their respective cost. As described in a subsequent section, this mode, with limited redesign, appears to be most appropriate for assessing the relative cost-effectiveness of individualized and standardized programs. During the exploratory phase of the study, 45 variables were found to discriminate in a statistically significant manner between successful and unsuccessful programs. In addition, over 30% of the variation in student performance, as measured by standardized norm-referenced testing, was explained by per-pupil costs of resources devoted to reading. The study identified a number of significant variables which were associated with individualized programs. These included: a) amount of teacher time devoted to diagnosis/prescription activities; b) occurrence of teacher redesign/development of performance objectives; c) diagnosing individual student needs as a major emphasis in pre-service training; d) amount of teacher time spent in planning activities; and other instructional management variables. As described in Exhibits I-3 and I-4, the cost associated with various staff activities and variables were also determined. The model used in Michigan, unlike others, attempted to determine the amount of resources, regardless of funding source, consumed by an individual child throughout the year. The average cost in successful Title I

Exhibit I-3

DOLLARS PER STUDENT FOR
TWENTY-FIVE HIGH SITES

HIGH SITES -- 25 AVERAGES RESOURCES	FUNCTIONS					
	COMP-ED READING	COMP-ED PLANNING	COMP-ED TRAINING	COMP-ED DECISION MAKING	COMP-ED ADMINI-STRATION	RESOURCE TOTAL
PERSONNEL						
District Comp-Ed Director	$ --	$ 17.84	$ 4.36	$ 12.80	$ 9.40	$ 44.40
Principal	--	15.08	7.80	13.92	9.20	46.00
Comp-Ed Teacher	193.60	53.52	10.84	30.60	--	288.56
Regular Teacher	93.84	92.00	10.60	49.80	--	246.24
Paraprofessional	47.20	3.24	.80	1.00	--	52.24
Reading Specialist	1.16	2.64	.68	2.24	.84	7.56
Other Classroom Staff	.24	--	--	2.12	--	2.36
CONSUMABLES						
Comp-Ed Books and AV Software	24.40	--	--	--	--	24.40
Regular Books and AV Software	9.88	--	--	--	--	9.88
EQUIPMENT						
Comp-Ed AV Equipment	2.76	--	--	--	--	2.76
Other Comp-Ed Instructional Equipment	4.28	--	--	--	--	4.28
Regular AV Equipment	--	--	--	--	--	--
Other Instruction Equipment	3.72	--	--	--	--	3.72
Comp-Ed Administration Equipment	--	--	--	--	.16	.16
MISCELLANEOUS						
Miscellaneous Comp-Ed Training Expenses		--	6.28	--	--	6.28
Miscellaneous Comp-Ed Administrative Expenses	--	--	--	--	3.16	3.16
FUNCTION TOTAL	$381.08	$184.32	$41.36	$112.48	$22.76	$742.00

Exhibit I-4

DOLLARS PER STUDENT FOR
TWENTY-THREE LOW SITES

HIGH SITES -- 25 AVERAGES	FUNCTIONS					
RESOURCES	COMP-ED READING	COMP-ED PLANNING	COMP-ED TRAINING	COMP-ED DECISION MAKING	COMP-ED ADMINI-STRATION	RESOURCE TOTAL
PERSONNEL						
District Comp-Ed Director	$ --	$ 16.83	$ 4.22	$17.48	$13.52	$ 52.05
Principal	--	12.09	7.70	19.00	13.57	52.36
Comp-Ed Teacher	135.96	17.78	2.83	11.78	--	168.35
Regular Teacher	85.57	65.48	5.04	34.74	--	190.83
Paraprofessional	70.26	.78	.30	1.22	--	72.56
Reading Specialist	6.04	3.13	1.57	2.00	--	12.74
Other Classroom Staff	--	2.83	--	.78	--	3.61
CONSUMABLES						
Comp-Ed Books and AV Software	13.04	--	--	--	--	13.04
Regular Books and AV Software	9.87	--	--	--	--	9.87
EQUIPMENT						
Comp-Ed AV Equipment	1.17	--	--	--	--	1.17
Other Comp-Ed Instructional Equipment	.22	--	--	--	--	.22
Regular AV Equipment	--	--	--	--	--	--
Other Instructional Equipment	3.00	--	--	--	--	3.00
Comp-Ed Administration Equipment	--	--	--	--	.39	.39
MISCELLANEOUS						
Miscellaneous Comp-Ed Training Expenses	--	--	4.09	--	--	4.09
Miscellaneous Comp-Ed Administrative Expenses	--	--	--	--	2.91	2.91
FUNCTION TOTAL	$325.13	$118.92	$25.75	$87.00	$30.39	$587.19

reading programs of $850 was significantly higher than conventional wisdom would indicate and suggests serious questioning of the $300 critical mass figure suggested in the HEW cost-effectiveness study of compensatory education (Lynn, 1972). During the cross-validation study presently underway and scheduled for completion in November 1975, analysis will determine the instructional program characteristics, if any, which are associated with successful programs. Such an analysis should provide greater insights regarding the relationship between activities inherent to individualized instruction and program outcomes. In addition, as described in a subsequent section, the model will also provide information on the trade-offs associated with various types of instructional programs. the COST-ED model, modified for application in Michigan, is unique in its capacity for determining trade-offs and conducting sensitivity analyses.

In summary, very little is known about the effectiveness and associated costs of factors contributing to the success of individualized systems. The concern expressed about the cost of individualized instruction (e.g., the RBS proposal) and the extremely low advertised costs of implementing individualized processes such as IGE (e.g., $10 per pupil) indicate a dire need to address this issue in the design.

C. Implementation of Individualized Programs

Existing procedures for assessing degrees of implementation are inadequate for the purposes of the study, thus requiring extensive design/development effort during the implementation phase.

While specific details regarding the issue of assessing degrees of implementation are addressed in Section III/Task 3 of this report, based upon an analysis of the SRI Follow Through instrument, the/I/D/E/A/ Implementation Kit, and various implementation instruments (either observation or survey) used to assess various learning systems, the following findings in this area appear to be germane: a) that attempts to develop single instruments which would apply generally across all programs, such as the Follow Through instruments (SRI) have resulted in findings of questionable validity; b) no existing instrument has been designed to assess degrees of implementation, comparing well-implemented individualized and standardized programs; c) existing observation and survey instruments to assess the degree of implementation of specific learning systems appear to be appropriate for redesign and use as a data collection instrument; and d) the implications for the analysis plan are significant indeed in that cross-program comparisons will be limited.

D. Parental Involvement

Due to its research void and inherent sensitivity, parental involvement can be treated as a progam variable only in a limited manner in this study. It should be addressed in detail in a separate study.

In reviewing the minimal existing literature on the impact of parental involvement in compensatory education (Stearns) and based on observations and discussions with project staff, the following findings, described in greater detail in Appendix 2, have been made: a) there exists a great void in research on parental involvement and its impact upon student performance in compensatory education, even through parental involvement is increasingly becoming a requirement in most comp ed programs; b) the types of parental involvement, in terms of compensatory education generally, include advisory functions through Parent Advisory Committees, as employees providing assistance to teachers in the classroom, as tutors or instructors in at-home programs or components, and as participants in the establishment of individual objectives; c) the quality of parental involvement in the above categories varies considerably, requiring the development and use of instruments which will capture qualitative differences and contextual situations. Based on the above findings, we are proposing a separate study, as described in Appendix 3.

In this section of our Report, we have described the background of individualized instruction in compensatory education programs and discussed the state of the research art as it relates to the purposes of this study.

References

Section I

Boardman, Gerald R. and Hudson, C. Cale. *Development of a Cost Analysis Model Which Schools May Use to Determine Budget Needs for Implementing Individually Guided Education (IGE).* Lincoln, Nebraska: Teachers College, University of Nebraska, 1973.

Brayn, Dennis. "An Evaluation of the Menominee Project MATH". Lansing, Michigan: Michigan Department of Education, 1974.

DeVault, M. Vere, et. al. *Descriptor for Individualized Instruction.* Madison, Wisconsin: Wisconsin Center for the Analysis of Individualized Instruction, 1973.

Edwards, Floyd H. *A Study of Affective Change in Elementary Schools Implementing Individually Guided Education.* Unpublished doctoral dissertation. University of North Carolina at Chapel Hill, 1972.

Gramenz, Gary W. *Relationship of Principal Leader Behavior and Organizational Structure of the IGE/MUS-E to I & R Unit Effectiveness.* Technical Report No. 320. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1974.

Gresso, Don W. *Individually Guided Education and Organizational Climate.* Unpublished doctoral dissertation. University of Missouri-Columbia, 1974.

Herrick, H. Scott. *The Relationship of Organizational Structure to Teacher Motivation in Multiunit and Non-Multiunit Elementary Schools.* Technical Report No. 322. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1974.

Kelley, Edgar A., Wood, Fred H., and Joekel, Ronald. *Teacher Perceptions of School Climate and the Implementation of Individually Guided Education (IGE).* Unpublished doctoral dissertation. University of Nebraska, 1973.

Lazich, Gilbert S. *The Effects of the Wisconsin Design for Reading Skills Development in K-3, Niles Michigan, 1971-73.* Hester, Pennsylvania: Wayne State University, 1974.

Lynn, Lawrence. *Cost Effectiveness of Compensatory Education.* United States Department of Health, Education and Welfare, 1972.

Mayfield, Irene R. A Comparative Study: Two Methods of Teaching Mathematics - Conventional and Individualized. Jackson, Mississippi, 1973.

Nelson, Richard G. An Analysis of the Relationship of the Multiunit School Organizational Structure and Individually Guided Education to the Learning Climate of Pupils. Technical Report No. 213. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1972.

Pellegrin, Roland J. Some Organizational Characteristics of Multiunit Schools. Working Paper No. 22. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1969.

Richardson, Edward R. A Study of the Changes in Role Perceptions and Role Behaviors of Principals in Individually Guided Education in Multiunit Elementary Schools. Unpublished doctoral dissertation. Auburn University, 1972.

Rossmiller, R., et. al., "Research and Evaluation Related to IGE", (Draft No. 4, July 1975). Madison, Wisconsin: Wisconsin Research & Development Center for Cognitive Learning.

Smith, Kenneth B. An Analysis of the Relationship Between Effectiveness of the Unit Leader in Conducting Unit Functions. Technical Report No. 318. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1974.

Stearns, Marian and Peterson, Susan. Parental Involvement in Compensatory Education. Palo Alto, California: Stanford Research Institute, 1973.

Tallmadge, Kasten. A Guide to Project Evaluation in Education. Resources Management Corporation, 1975.

Walter, James E. The Relationship of Organizational Structure to Organizational Adaptiveness in Elementary Schools. Technical Report No. 276. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1973.

SECTION II

GENERAL APPROACH/STUDY DESIGN

In this section we discuss the general approach which we are proposing, the major design issues and technical approaches taken in each of the tasks outlined in the original RFP and our general philosophy which permeates the proposed evaluation design. Where appropriate, to reduce redundancy, we reference prior sections both in this report and in our initial Proposal, in discussing justifications and rationales.

MAJOR STUDY OBJECTIVES

Based upon the general objectives specified in the RFP and in light of the findings which arose during the design phase, we are proposing the following major study objectives:

1. To determine the differential effects and effectiveness of well-implemented individualized vs. well-implemented standardized math and reading programs which provide instruction for compensatory education students at the elementary level. More specifically:
    - to evaluate the relative impact on cognitive growth in math and reading;
    - to compare the impact upon student and teacher attitudes and student outcomes in the affective domain; and
    - to identify the relationships among individualization, classroom environment, and school climate, identifying and analyzing possible reasons for these relationships.
2. To determine the relative cost-effectiveness of the process dimensions of individualized instruction.
3. To determine the degrees to which potentially well-implemented programs are implemented during the period of observation and in turn, the relationship between degrees of implementation and program outcomes within and among programs.

SAMPLING PLAN

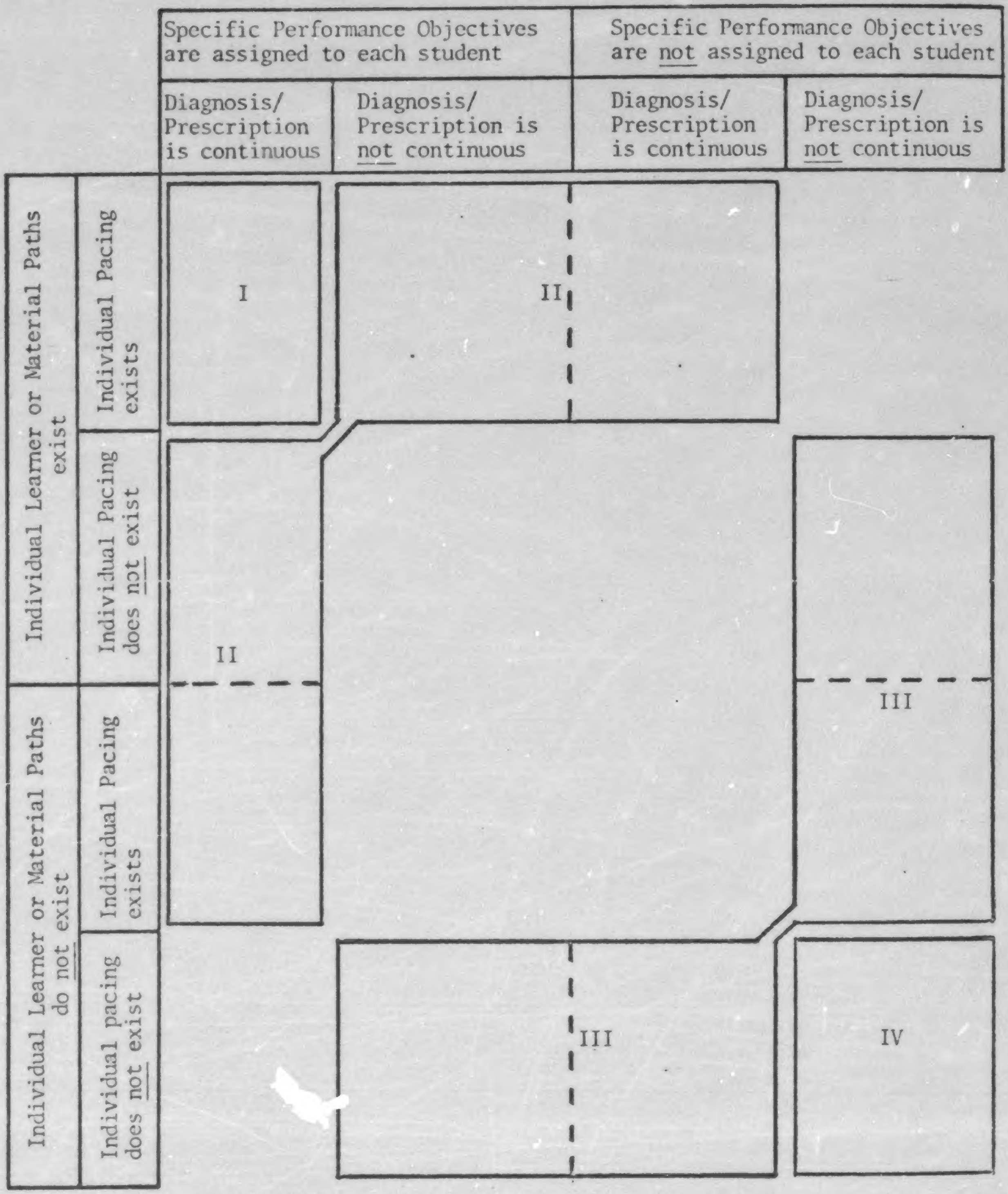
The sampling design which we are proposing requires data from approximately 800 classrooms with equal or proportional representation in each of the four categories (I, II, III, and IV) illustrated in Exhibit II-1. Equal or proportional representation within math/reading and by mainstream/pull-out classroom settings will be maintained. An equal number of classrooms from grades 2 and 3 are specified at each substrata of this framework. Potential candidate sites have been identified (or will have been identified) through District Survey I in accordance with its initial data collection effort and through the efforts of the project team, the latter including selection from documented and validated well-implemented programs, nominations of well-implemented programs from publishers and/or model developers, and TURNKEY and other documentation sources. During the design phase, "potential" candidates were verified to be "likely" candidates through telephone conversations and extensive documentation review, to ensure the feasibility of the sampling plan. During the implementation phase, further verification of the sampling plan through questionnaires and on-site visitations will be required prior to final selection of participants.

VARIABLES

Based upon a review of relevant literature and research findings, including studies conducted by TURNKEY, and in accordance with our sampling plan, we have identified three categories of independent variables in the proposed "planned variation" study: a) type of program, according to the degree of individualization; b) instructional setting (e.g., mainstream vs.

Exhibit II-1

BASIC SAMPLING FRAMEWORK


Specific Performance Objectives are assigned to each student
Specific Performance Objectives are not assigned to each student
Diagnosis/ Prescription is continuous
Diagnosis/ Prescription is not continuous
Diagnosis/ Prescription is continuous
Diagnosis/ Prescription is not continuous
Individual Learner or Material Paths exist
Individual Pacing exists
Individual Pacing does not exist
Individual Learner or Material Paths do not exist
Individual Pacing exists
Individual pacing does not exist
I
II
II
III
III
IV

pull-out); and c) subject area (i.e., math or reading). In addition to a number of control variables, including: a) student achievement in math and reading; b) classroom environment and climatic variables; and c) affective variables relating to students and teachers, including satisfaction. In addition, we have identified a number of implementation variables which will also be used as program variables to relate to various outcome variables.

INSTRUMENTATION

The instruments to be utilized in the proposed design include: a) a limited number of existing instruments; b) existng instruments which will necessarily require modification; and c) newly-developed instruments.

In order to assess student outcome measures in math and reading, we are proposing to utilize the Stanford Achievement Test (Primary I, II, and III) at grades 2 and 3. If after program selection it is determined that the objectives and content of the programs differ significantly, we propose to use selected items from the SAT as objective-referenced or criterion-referenced tests to ensure program fairness. A unique set of procedures based on the work of Klein is proposed.

For interview purposes, we are proposing to modify the instruments for interviewing staff used in the Michigan study of compensatory education conducted by TURNKEY. For the purposes of measuring student outcomes in the affective domain, we are proposing that the revised Piers-Harris Children's Self-Concept Scale (recently validated in an inner-city school for grades 1-3 by the Purdue Education Research Center under a USOE contract) be used along with the Attitude Towards School to assess student attitudes towards school; the Organizational Climate Index will be used to measure teachers' assessment of classroom environment. Individual instruments following a uniform format are proposed to assess degrees of implementation in the form of checklists and/or observer verification forms. As an option, we have included an instrument which could be used for assessing implemention of the dimensions of individualization, if this instrument is found to be feasible during field testing.

DATA COLLECTION

The data collection plan which we proposed for project implementation consists of three types of data collection procedures:

- Classroom observations, (including verification);
- Staff interviews (both structured and open-ended); and
- testing.

Observations of compensatory education classrooms will occur periodically during the 1976-77 school year in all classrooms. Interviews of principals, teachers, aides, and other appropriate staff members will be conducted early in the school year and followed-up with parallel verification interviews in the spring of 1977. Achievement and other tests will be administered as early as possible durng the 1976-77 school year as a pre-test and again later in the spring as a post-test, as prescribed in the SAT Manual.

The data collection itself will be supervised and conducted by local personnel and will be coordinated by a regional structure of core staff members. These data collectors will be experienced professionals who will have familiarity with local policies, procedures, and personnel.

Through the use of local personnel rather than outside data collectors, the twin problems of data confidentiality and site cooperation may be more readily addressed.

DATA REDUCTION AND ANALYSIS

Data obtained throughout the 1976-77 school year from pre-test administration and from periodic observation of program operation and characteristics will be reduced for analysis upon collection. Procedures for the speedy reduction of the post-test results and other late school year data will be carefully refined during this school year to ensure that, once the data is obtained in May 1977, all necessary data will be reduced and analyzed in time to report the study's results in July 1977. Since TURNKEY experience has indicated the extremely critical nature of these data reduction tasks, sufficient resources will be allocated to this activity in our proposed design to ensure their timely completion.

Once all data has been obtained, the carefully planned data analysis effort will proceed. This effort will address all research questions to be covered by the study in a specific sequence. These questions are hierarchical in nature for the proposed study, reflecting the interaction effects nature of our semi-factorial study design. The techniques to be applied in our proposed series of control contrast range from simple comparison of means or frequencies via tests or contingency table methods to more sophisticated methods such as ANOVA discriminant function analysis.

UNDERLYING PHILOSOPHY OF EVALUATION

Reflected in the above general approach and implicit in our discussion of how we propose to conduct specific tasks is an underlying philosophy about evaluation generally and about federally-sponsored evaluation studies/experiments in education in particular. While we are cognizant of recent literature regarding studies of policy research in education, one's own experience as a participant and/or outside observer of large-scale study efforts similar to this one necessarily tends to influence one's approach. A brief understanding of this philosophy will perhaps put some of the subsequent discussion in a propoer perspective.

First, we believe that the ultimate purpose of evaluation is to provide useful and timely information to decision-makers for the purpose of improving programs, more so than proving whether or not a program works. The "users" of information resulting from this study will be Congress as they formulate education policy and, ultimately, local education agencies as they have the prime responsibility for ensuring improvement of the operations of compensatory education programs. Proving that a particular program works is less interesting than determining why and under what conditions. Moreover, when the perceived purpose of evaluation is to "improve" than than "prove", the negative connotations sometimes directed toward recent federal evaluations can be minimized.

Second, while we realize that social science research in general and, specifically research related to education, is highly challenging, we also feel strongly that the federal role in evaluation presents an awesome responsibility. We applaud attempts at the federal level to improve the state of the art in social science research. However, as the federal government experiments with new evaluation and experimental techniques one has to realize that education officials at the SEA and LEA levels are more concerned with "reported results" than the advent of new techniques.

In the recent past allegations have been made that one of the reasons why "commonly accepted successful programs" did not fare favorably in national evaluations was due to the inadequacy of the evaluation designs specifically and the inability of the Government (for inter-governmental, political and other reasons) to conduct experiments/evaluations effectively. In our proposed design we have attempted to strike a balance between proposing techniques and instruments which we feel are technically sound and appropriate (although they may be non-traditional), while at the same time seeking to minimize the probability of questionable findings due to research flaws.

Third, we feel that large-scale federal evaluations and study designs should be flexible and accomodating rather than rigid and parochial, taking into account the political, social, and economic milieu in which education actually occurs in this country. For example, while true and pure experimental designs may be relevant for disciplinary research, the feasibility of applying such designs in policy research of this nature may create the need to gravitate towards "quasi-experimental" designs. Similarly, in data collection one might have to opt for instruments and procedures based upon a trade-off between what is technically possible and what can be accomodated by the LEAs and programs being studied to ensure a modicum of cooperation necessary for the completion of any study of this kind. We feel this is particularly true in this study where most of the programs studied will have been developed and/or operated through resources which, for the most part, are locally generated and/or allow local discretion.

Fourth, federal evaluation techniques and research designs should be appropriate and should ensure "fairness". Too often the criteria used to judge the effectiveness of programs at the LEA level differ significantly in terms of priority rankings from those at the federal level. Care should be taken to ensure that success criteria are generally accepted and where divergent criteria exists, the design should accomodate a number of them. Our proposed design reflects this consideration. Moreover, evaluation techniques should be program-fair, assessing the progress achieved in programs in accordance with intended purposes. For this reason, we have taken great care and suggest a considerable amount of time to be devoted to ensuring that instruments, particularly implementation instruments, do in fact assess what is supposed to happen at the classroom level. While many tenets of the above philosophy have been espoused before, we propose that the implementation contractor use the above principles as a guide in implementing the proposed design.

SECTION III
DESIGN ISSUES/TASK APPROACHES

In this section of our report we outline the specific design issues and approaches to the first five tasks specified in the RFP. Note that the tasks we refer to here are listed as subtasks of the overall study design task on page 20 of Enclosure I in the RFP. We have adopted the notation of nine "tasks" rather than nine "subtasks" within an overall design "task" simply for convenience. This section of our report concentrates on the five "design" tasks (as opposed to "data collection," which are discussed in Section IV).

TASK 1 -- DEFINE PROGRAM VARIABLES

Design Issues

The sampling, control, and additional descriptive variables of interest to the study must be carefully and rigorously determined since these definitions will affect all other aspects of the study. Sampling procedures, interview and observation schedules, the data analysis plan, and the conduct of data collection efforts all depend, for their framework, upon the definitions of program variables arrived at in this task. The ability of NIE and the implementation contractor to assert that they assessed the effectiveness of individualized instruction in reading and mathematics programs for compensatory students will depend in large part upon the quality of these definitions. Ultimately, Congressional use of the results will also depend upon the utility of these definitions and the quality of the proposed design.

As indicated in the RFP, a crucial aspect in defining the program variables will be the identification of operational criteria of what is an "individualized" program or a "standardized" program. The TURNKEY team recognizes that the criteria, as listed in the RFP, are basic:

- individualized programs will have stated or written performance objectives prior to implementation of the instructional strategies relied upon to meet these objectives;
- individualized programs will rely upon diagnostic procedures for assessing the instructional needs of each individual student relative to the stated objectives;

- individual programs will provide instruction to each individual child that is a prescriptive reflection of the diagnosis of his needs relative to the stated objectives;
- individualized programs will rely upon progress tests in order to assess mastery at regular intervals during the program; and
- individualized programs will be characterized by individually paced instruction and performance-based student progress and instruction tied to the student's results on progress tests.

This list of characteristics must, however, be operationalized further for a number of reasons.

At this time a large number (to be increased in 1976 school year because of new Title I guidelines) of compensatory education programs operate with a set of stated or written performance objectives, although a sizeable variation may exist among the quality and appropriateness of these objectives. The results of our recent study of compensatory education practices in Michigan indicate that 71% of the compensatory education teachers sampled from the most "successful" districts stated that they provided reading instruction that was directly related to a set of written performance objectives. Similar teachers for the "unsuccessful" districts indicated that 63% of their number also provided instruction related to such objectives. Moreover, where these objectives apparently existed, 80% of the respondents claimed that the objectives were tied to individual student performance rather than group performance -- for both successful and unsuccessful programs. In the Michigan study design, programs were not selected on a scale of individualized vs. standardized, but rather on a scale of program success. From this sample of Michigan programs, however, it appears that nearly 70% of Michigan compensatory education programs were operating with such written objectives. Since it is unlikely that those districts, which rely upon written objectives, all provide completely individualized programs, it is apparent that while the existence of such objectives is necessary, it is certainly not a sufficient condition to allow a program to be labelled "individualized."

Further, it is also true that many teachers rely upon some type of diagnostic testing, especially for compensatory programs where Federal guidelines set forth such a procedure. The tests relied upon for diagnostic testing may be specially developed or standardized or may even be subjective judgments; but, whatever their form, such tests are -- like written objectives -- so common today that their existence is only a potential indicator of individualization, not a guarantee.

Similar concerns may be raised relative to prescription-based instruction and the use of progress tests to assess mastery. However, progress tests may not be as commonly used as diagnoses/prescription and may, therefore, be a better measure of individualization than the others so far discussed.

Even the key element of individual-pacing, where instruction is tied to measured student progress, must be examined critically to assess the uniqueness of these characteristics to individualized instruction. In the Michigan study, teachers in the successful sites were asked what percent of their students' reading time was spent in individually-paced activities; their average response was approximately 85%. Similar teacher in the unsuccessful sites responded with an average of about 74%. Thus, the aspect of individual pacing was fairly common to all programs studied, indicating that individual pacing (or teachers' perceptions of individual pacing) is not necessarily synonomous with individualized instruction or that individualized instruction was not a unique characteristic of successful programs.

Ultimately, it might be that the key characteristics of an educational program that would allow that program to be incontrovertibly labelled "individualized" is the individualization of the program's response (in terms of material, content, techniques, etc.) to the results of the progress checks. For instance, if all learners in a program use the same materials but some are allowed to use them at a slower pace than others, then it is quite possible -- even with all of the other accoutrements of individualization in force -- that this program is only a time-eased version of our traditional programs. But in order for a truly individualized instruction plan or set of instructional events to be carried out in response to each individual child's progress test results, then each of the preceding elements (objectives, diagnosis, prescription, progress evaluation) must be part of the overall program. The potential for self-pacing is inherent in the existence of multi-paths for the instructional plan to follow. Thus, although the existence of these multi-paths might be necessary and sufficient in itself to justify a label of "individualized", the existence of these other aspects is necessary to validate any claims of effective individualization.

While this discussion of compensatory education as actually practiced in classrooms was intended to illustrate the difficulty in contrasting programs labelled "individualized" with those labelled "standardized", by defining a "standardized" program as one which contains none of the elements just discussed one would probably find only a very few such programs within the universe of Federally-supported compensatory education programs. Given the moderating effects of LEA and SEA regulations and traditions (e.g., over 14 SEAs mandate various types of assessment objectives) defining an "individualized" program as one which completely embodies all of these aspects to a maximum degree would also result in the identification of only a very few such programs. Even if such extreme cases could be found, it is clear that policy research (considering Congress as its prime audience) should not be greatly concerned with such unrepresentative cases.

In our Proposal (p. III-7) we suggested a two-staged activity to operationally define individualized and standardized programs during the design phase. To summarize, five questions relating to: sequence, pacing, use of individual progress checks, use of diagnosis and prescription, and documentation of written performance objectives would be asked and scaled on a continuum indicating the degree of individualization (e.g., number of Is vs. S'S). The second step as originally proposed would include a review of documentation to be gathered during the implementation phase of the contract to validate the initial responses. As described previously, the project team collected documentation on a large number of projects and conducted a second stage activity for many sites during the design phase. In addition, a limited number of observations were conducted in conjunction with other projects and activities involving members of the project team. The net results of this "field testing" of the approach resulted in the final proposed definitional framework which has been included in our sample design.

The definitional matrix will be used to classify whether a program meets the following characteristics:

"PO"	Whether specific written (documented) performance objectives are assigned to individual students. Objectives would also include stated/commonly known criterion or proficiency levels indicating mastery (paper and pencil or performance based). "Givens" and "time periods" would be considered desirable but not necessary.
"D/P"	Whether initial placement for each student in the program's materials is based on a diagnostic test and a specific process is subsequently followed for diagnosing student needs and assigning prescriptive materials or exercises on a continuing basis. By 1976, all Title I schools will be required by regulation to administer "initial" or "placement" diagnostic tests to all students; hence, the emphasis on continuous diagnosis.
"ALP"	Whether uniquely prescribed individual learner paths through the relevant program materials, related to performance on prior activities, are followed by each student. Preliminary review of project documentation and discussions with LEA staff indicated that in a limited number of programs which were always described as being very "individualized", in all instances individual students participated as part of a small group for relatively short periods of time in conducting specific exercises or drills. Hence, an individual learning path which includes some participation in small groups or specific activities would be allowed. This dimension is, in fact, an extension of the prescription characteristic.
"PACE"	Whether or not the amount of time required for any student to master a specific performance objective or to complete a given portion of a program is determined individually for or by that student and varies from student to student.

Program Variables

The RFP asked the design contractor to identify program operating characteristics of interest, control, and descriptive variables. In our Proposal we discussed a number of categories of descriptive variables used by TURNKEY in its study of compensatory education programs in Michigan. Specifically, nine groups of variables were hypothesized as acting together in their impact on program effectiveness. Through the application of various statistical techniques, 45 variables were found to discriminate statistically between successful and unsuccessful programs. This study is continuing and, at present time, is in a cross validation phase with expected results in November 1975. The specific variables which were significant during the exploratory phase of the study are included below as part of the overall list of variables and specifically in Appendix 4. In addition, as noted in the prior section of this document an extensive review of recent evaluations relating to this study was conducted from which additional program variables were identified. The results of our efforts during the design phase thus far are listed below in terms of the various types of categories of variables. It should be noted that implementation variables are discussed in Task 3.

The variables to be studied by means of the proposed study design fall into three major categories: Independent, Moderator, and Dependent. Each of these categories is described below and the specific items to be studied within each item listed and the rationale behind the inclusion of the item are also shown.

A. Independent Variables

The variables included within this category are those program factors whose impact we wish to investigate directly by incorporating them into the (stratified) sampling plan for the study. We accomplish this by means of a specific sampling plan which follows an analytical structure designed to address these variables explicitly. There may be other program factors whose impact we would wish to study, but only those reflected in our sampling design can be viewed as independent variables. The program factors included in this category are listed in Exhibit III-1 along with the suggested data collection method and the rationale for the inclusion of each item.

B. Moderator Variables

The variables included within this category are those program or student factors whose impact we wish to control by holding their values constant or to study as covariables with our dependent variables. The first of these two types of moderator variables is referred to as Constants, and the second as Covariables. Each of these two groups is described in further detail below. Although the covariables are not reflected directly in the sampling plan, their impact on the dependent variables can be assessed statistically by incorporating these variables as covariates in analyses of covariance.

1. Constants

These are program or student characteristics whose impact we wish to hold constant for all analyses which are conducted as part

EXHIBIT III-1

INDEPENDENT VARIABLES

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Type of program as categorized via definitional matrix (see Task 2) i.e., Whether a program is very individualized (Group I) individualized (Group II) standardized (Group III) very standardized (Group IV)	Sampling data (see Sampling Plan) verified during on-site visit to specific classrooms plus defined in more detail via on-site observations in specific classrooms. Observations/Survey items to include: • student record folders to note time required to complete units or materials; • identification of fraction of core materials which are used by some but not by all students; • the diagnosis/prescription process; • identification of objectives associated with specific students including the extent to which students can describe the objective currently being addressed.		X		X
Instructional setting of specific comp ed instruction; specifically whether the program operates in a mainstream or a pull-out fashion.	Sampling data (see Sampling Plan) verified during on-site visits to specific study schools. Data required would include a description of what portion of the instructional program is provided in what setting for the comp ed students		X		X
Subject area: specifically reading and math.	Sampling data (see Sampling Plan).		X		

of this design. In reality, these constants may have an impact on the dependent variables; but for the reasons detailed in Exhibit III-2, we choose to view their impact as being beyond the scope of this study. These reasons are listed in Exhibit III-2 along with other information related to each item in this category in a manner similar to Exhibit III-1.

2. Covariables

These are program factors or student characteristics whose impact we wish to study (as is the case for the independent variables). However, we have not specifically reflected a balanced and ordered sample of relevant levels of each of these blocking variables in our sampling plan, a process which was specifically followed for the independent variables. Thus, the covariables are measured in a manner similar to the independent variables, and their impact may be studied only to the extent that the data resulting from the sampling design provides a reasonably representative distribution of values for the covariables. The program factors and student characteristics included in this category are listed in Exhibit III-3 along with the suggested data collection method and the rationale for the inclusion of each item.

C. Dependent Variables

The variables included within this category are those student or school characteristics we are using as the measures of effectivenes in the proposed study design. These measures fall into two major areas:

EXHIBIT III-2
Page 1 of 2

CONSTANTS

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Grade level -- all programs studied will be operational at grades 2 and 3 and classrooms for each of these levels will be equally represented. While grade level comparisons between these two levels will be possible, it is felt that by studying these contiguous levels we can best understand the impact of individualization -- a concept in which grade level becomes less meaningful beyond administrative needs.	Sampling data (see Sampling Plan), verified on-site.		X		X
All students will be "compensatory education students", i.e., either Title I or Title I eligible (it is felt that this wider definition is needed to identify sufficient Group III and IV programs).	Sampling data (see Sampling Plan), verified on-site.	X	X		X
All programs will have been operational for at least the two previous years in the LEA, and at least the previous year in the subject school (for stability purposes and to help ensure "well implemented" programs).	Sampling data (see Sampling Plan), verified on-site.	X	X	X	X
Subject students as much as possible be all native speakers of English at home (to minimize testing problems).	Sampling data (see Sampling Plan), verified on-site.	X	X		X

EXHIBIT III-2
Page 2 of 2

CONSTANTS (Continued)

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
All programs will be "well-implemented", whether individualized or not (design constraint).	Sampling data (see Sampling Plan), specifically the implementation/effectiveness question, the stability items (programs, teachers, students), the documentation items. During the study, implementation to be addressed via observation of actual definitional matrix category vs. initially assessed category, controls placed on testing scoring procedures to reflect actual instructional content in each program, and observer judgment of the degree to which the program seems well run according to a list of general guidelines for well-run programs that would be developed by the study contractor.	X	X	X	X

EXHIBIT III-3

COVARIABLES

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Number of years that the subject student (Ss) have been in the program; specifically one or more than one.	Sampling data where possible verified during on-site visits; but mainly through survey data gathered during the study year.		X		
Gender of Ss.	Survey data gathered during study year.		X		
Parental involvement in program.	Observation of parental classroom activities, if any, during the study year; mainly through survey data gathered during the study year.	X	X		
Class size.	Survey and observation data gathered during the study year.	X	X		
Cost of the resources consumed in the program, both during instruction and in support activities.	Survey data from TURNKEY's cost-effectiveness instruments (see Dependent Variables - Process below).	X	X		
Pre-test scores (used to form residualized gain scores).	Pre-test score results for sample Ss.		X		
Presence/absence of relevant instruction in content area addressed by test.	Teacher survey; work or student progress test samples; examination of existing performance objectives.		X		X

outcome variables and other associated variables. Under the area of outcome variables, measures in both the academic (cognitive growth) and the affective areas are included. Under the area of other associated variables, measures in the area of classroom environment, as well as a number of factors to be studied post hoc (i.e., factors which are not reflected in the sampling design) are included. Each of these areas is described in further detail below:

1. Academic Variables

   These are measures which encompass cognitive growth in the basic skills of reading and math, as well as other such school-based measures of student performance as attendance and behavior patterns. Exhibit III-4 lists these factors.

2. Affective Variables

   These are measures which encompass the affective or attitudinal domain of student and school characteristics. Exhibit III-5 lists these factors.

3. Classroom Environment Variables

   These are measures which encompass general items related to classroom environment as well as specific factors which we wish to study after the fact. Exhibit III-6 lists these factors. Appendix 4 lists the other factors suggested for study as part of this category of variables. The items included in Appendix 4 are derived from TURNKEY's existing process evaluation instruments, which have been applied in numerous program studies throughout the

EXHIBIT III-4

ACADEMIC VARIABLES

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Residualized achievement gain scores (with pre-test co-variable effects removed).	Scores from standardized test - or sub scores (perhaps specially defined) from such tests, matched in coverage to the content areas covered in instruction.	X	X		X
Mastery of objectives.	Scores from criterion-referenced tests administered for this study, matched in coverage to the content areas covered in instruction (post-test only).	X			X

EXHIBIT III-5

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Student self-concept: feel self school self behaving self	Existing instruments including: Self Concept Ability Scale Piers-Harris Children's Self Concept Scale*		X		
Student's attitude toward: school work teacher school in general	Existing instruments including: Attitude Toward School* My Class Inventory School Attitude Test: Oral Form School Sentiment Index		X		
Attendance	Survey of school records	X	X		X
Disciplinary referrals	Survey of school records, coupled with a review of local disciplinary procedures.	X	X		X

*Instrument which is recommended for use.

EXHIBIT III-6

CLASSROOM ENVIRONMENT VARIABLES

ITEM	SUGGESTED DATA COLLECTION METHOD	RATIONALE			
		PRACTICALITY	DESIGN REQUIRED	STABILITY	IMPLEMENTATION
Staff assessment of classroom environment: intellectual climate achievement standards organize effectiveness supportiveness orderliness impulse control	Existing instruments including: Class Activity Questionnaire Learning Environment Inventory Minnesota Teacher Attitude Inventory Organizational Climate Descriptive Questionnaire Organizational Climate Index* Purdue Teacher Opinionnaire Social Climate Scale		X		X
Factors related to classroom and program process management data including: cost effectiveness staff involvement in program development program coordination in-service training (See Appendix for a full list of relevant variables.)	Existing instruments: TURNKEY's Process Evaluation Instruments and Observation items to be developed from subareas of these TURNKEY Instruments				

*Instrument which is recommended for use.

country. Listed in Exhibit III-6 for illustrative purposes are a number of process aspects found to be critical for effective programs in TURNKEY's work to date.

The analysis of the relationships between the previously listed items and the other factors suggested here is recognized as being one that could result in the identification of many apparently significant and important variables, if only because the number of items listed in Appendix 4 is quite large and a small percent (such as a chance rate of 5%) of a very large number is still a large number. Thus, care must be exercised in stating the implication of any results in the area. It is also recognized that such an analysis, even if identifying a truly significant number of important variables, would at best only form the basis for subsequent cross-validation studies that would control for the identified factors in a more systematic manner. However, since TURNKEY has already developed results from applications of the TURNKEY process evaluation instruments in a sizeable number of programs, it would be possible to use the existing TURNKEY data to predict the results from the NIE data base. Thus, the data even for this one year of study (1976-77) could serve to cross-validate the TURNKEY results to date by evaluating the success of each of these predictions.

Note that in Exhibit III-6 we have indicated that some of the items contained in the TURNKEY survey instruments mentioned there will be developed into an observation format. The observations obtained in this format would be collected along with the program implementation observations (see Task 3) and used to verify survey responses in these subareas obtained during early data collection interviews.

TASK 2 --DEFINE SAMPLING PROCEDURES

The procedures to be utilized in the determination of the sample of districts, schools, and classrooms to be included in the study fall into four principal subareas:

- the determination of the sampling frame/definitional matrix;
- the development of sampling criteria (i.e., for both degree of individualization and degree of implementation);
- the sample selection procedures; and
- final sample dimensions.

As was indicated in our Proposal, the selection of programs must include considerations of both degree of individualization and degree of implementation. The first of these considerations, would be estimated for any given program on a priori basis by a review of existing program documentation and tentative confirmation. The second consideration requires thorough analysis of program documentation and verification of activities (see Task 3). Our study design includes provisions for treating programs which, upon closer scrutiny during project implementation, are found to be less "well-implemented" than originally estimated.

In this section, we describe the recommendations and rationales for each of the four subareas which comprise the overall sample selection process.

The Sampling Frame/Definitional Matrix

The framework within which both the sampling and analysis processes must fit is determined by three dimensions: grade level, instructional setting, and type of program (degree of individualization).

A. Grade Level

The proposed study is to consider programs operated for comp ed students at grade levels K through 4. We recommend that only grades 2 and 3 be specifically examined within any program studied. We make this recommendation for the following reasons:

- It seems unnecessary to study programs at all five grade levels (e.g., for reasons used in the SRI Follow Through Study); yet the intended study must be beneficial at the lower elementary level. A subset of two of these five grade levels, properly selected, would suffice for this purpose.
- Any national study which ultimately must entail the use, in some manner, of standardized norm-referenced achievement tests should not rely upon present tests available at the K-1 levels. At these levels, the tests that do exist are more oriented to the concept of "readiness" than to that of achievement and require extensive 1:1 oral administration time. Thus, given limitations of using such tests, the study would then focus on grades 2-4.
- Since the basic notion of an individualized program is antithetical to the use of grade level labels, any grade level combinations used should be contiguous (i.e., grades 2 and 3 or grades 3 and 4 rather than grades 2 and 4) in order to reflect the contiguous nature of the instructional development within individualized programs. This constraint is desirable because slower students in the higher of the two levels might be using the lower level's materials and vice versa for faster students in the lower grade level. Moreover, content and sequencing vary considerably in individualized programs.
- The combination of grades 2 and 3 is preferred over the combination of grades 3 and 4 because the former is felt to be more representative of the K-4 range overall. Additionally, it may be that instructional programs at grade 4 are different enough in concept from those at grade 3 to make the 3-4 combination less contiguous than the 2-3 combination.

It is also suggested that for the purposes of analysis (see Task 5), all results for grades 2 and 3 should be combined, both for the individualized programs and for the standardized programs. Separate analyses by grade level could always be generated if a comparison of results by a category such as the nominal grade level associated with a given program for administrative purposes is considered desirable.

For sampling purposes, a constraint of having to include about the same number of nominally second grade classrooms and nominally third grade classrooms in each sampling cell would be necessary to reflect the grade level selection and discussion presented here.

B. Instructional Setting

The unit of analysis suggested for this study is the classroom. However, the special nature of the programs to be studied requires a careful description of the "classroom" to be studied. The concept of individualization suggests that important differences in process and/or educational practice may exist among different teachers within a given "individualized" program. Thus, many important process differences may be lost in an analysis which studies a unit larger than one tied to the activities of a specific teacher. It is also recognized that differences may exist among students served by the same teacher in an "individualized" classroom in terms of process applied or service received. However, since programs of general interest to the audience of this study are unlikely to have been implemented for only an individual student it seems inappropriate to limit the unit of analysis to an individual student, however incongruous that position might seem in a study of individualized instruction. Individualization within a classroom is a process controlled by the teacher, and it is the interaction of the teacher with the students in that classroom that should be an objective of analysis.

Having stated the case for the classroom as the study's unit of analysis, an additional complexity must be addressed. Since the study is to include programs of both the mainstream and the pull-out types, this issue must be considered in determining the actual unit of analysis. This consideration begins by understanding the basic nature of the comp ed reading or math programs that form the overall universe for this study, a universe from which well-implemented individualized and well-implemented standardized programs are to be drawn and compared on the basis of their effectiveness. The comp ed student receives the benefit of special funding provided by Congress (or other legislative bodies at the state and/or local level). This special funding is reflected in the comp ed school by the presence of specially hired personnel (e.g., teachers, specialists, paraprofessionals) and/or specially purchased materials or equipment.

The impact of these special personnel and/or materials is felt through interactions of these staff and/or materials and the comp ed students within these schools (not all students within comp ed schools are comp ed students) during instructional time periods. This instruction might take place within a portion of the regular classroom with the special comp ed instructor working with the comp ed students while the regular teacher in that room works with the non-comp ed students. Or it might take place within this same classroom with the regular teacher providing all instruction but with the assistance of supplementary materials for the comp ed students. Either of these operational structures would be considered a "mainstream" program.

Another means for conducting the comp ed instruction would be use of a classroom set aside especially as the comp ed learning center. In this case, the comp ed students would leave their regular classroom and receive instruction from the comp ed personnel in the center. This program would be of the "pull-out" type.

It is important to note that the mere existence of a center within a comp ed building does not label the comp ed program in that building as a pull-out type of program. The key issue is whether the comp ed students receive their comp ed instruction while in the same setting as their non-comp ed classmates or whether they are physically separated into different rooms for this instruction. If the non-comp ed students use both a base classroom plus the center for reading and math instruction, and the comp ed students in this class are always in the same room as their classmates even though the two groups receive different instructional programs in at least one of these settings, this program would be considered a mainstream type.

Further, while the programs for this study will be classified and selected based upon the nature of the comp ed instructional activities, comp ed students may participate in non-comp ed instructional activities as well as comp ed activities. It might occur, for instance, that a highly individualized comp ed program would operate within a "regular" setting that is highly standardized. The importance of the interaction between the regular and comp ed programs may depend heavily upon whether the comp ed program is of the mainstream or the pull-out variety.

It seems neither feasible nor desirable to build into the sampling strata all the possible interactions between the type of instruction in comp ed program and the type of instruction in the regular program. All such interface considerations (e.g., total hours of regular plus comp ed instruction vs. the hours in the comp ed portion of this total) which cannot be built into the sampling design will be considered in the overall study analysis as best they can, though it is recognized that only those process differences reflected in the sampling plan can be investigated in a exhaustive manner.

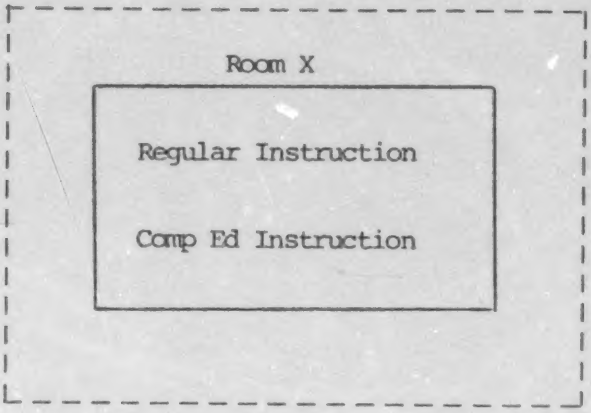
However, it is critical to the study that the instructional programs studied, which will be initially classified on the basis of only the specific comp ed instructional activities, be defined for analysis purposes as all instruction provided to the comp ed student whether by or in the specific comp ed program or by the regular staff in the regular classroom context.

Thus, the "classroom" to be considered the unit of analysis is a special one. Exhibit III-7 represents this "classroom" schematically for mainstream programs. This "classroom" -- indicated within the dashed boxes -- may indeed be one physical room (A) or may even be two rooms (B) so long as the comp ed students are not physically separated from their classmates during comp ed instruction. The program which we wish to investigate then would be the reading or math activities of the comp ed students provided both by the comp ed program and by the regular teacher as part of the general program.

Exhibit III-7

EXAMPLE STUDY "CLASSROOMS" FOR MAINSTREAM PROGRAMS

STUDY "CLASSROOM" A


Room X
Regular Instruction
Comp Ed Instruction

STUDY "CLASSROOM" B

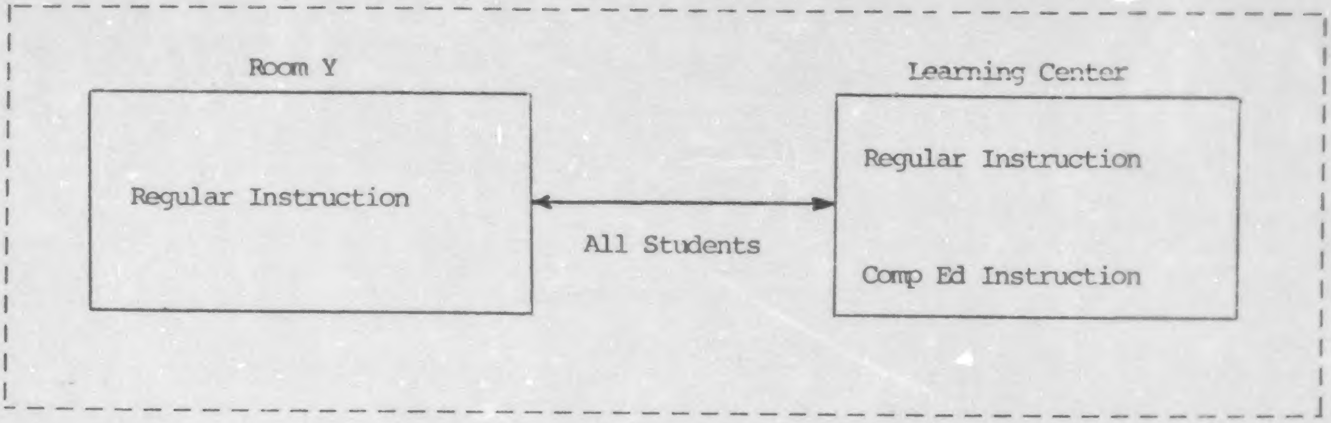

Room Y
Regular Instruction
All Students
Learning Center
Regular Instruction
Comp Ed Instruction

Exhibit III-8 represents the study classroom for pull-out program. The program which we wish to investigate then would be the reading or math activities of the comp ed students served by both the comp ed program and by the regular teacher in the classroom from which the comp ed student has been pulled. The specific rooms to be encompassed in our study "classroom" in this case would depend on the specific students in the center whose results are to be sampled.

C. Classifications of Programs

As discussed in Task 1, each program studied will be classified in a dichotomous (Yes/No) manner on each of the following four issues:

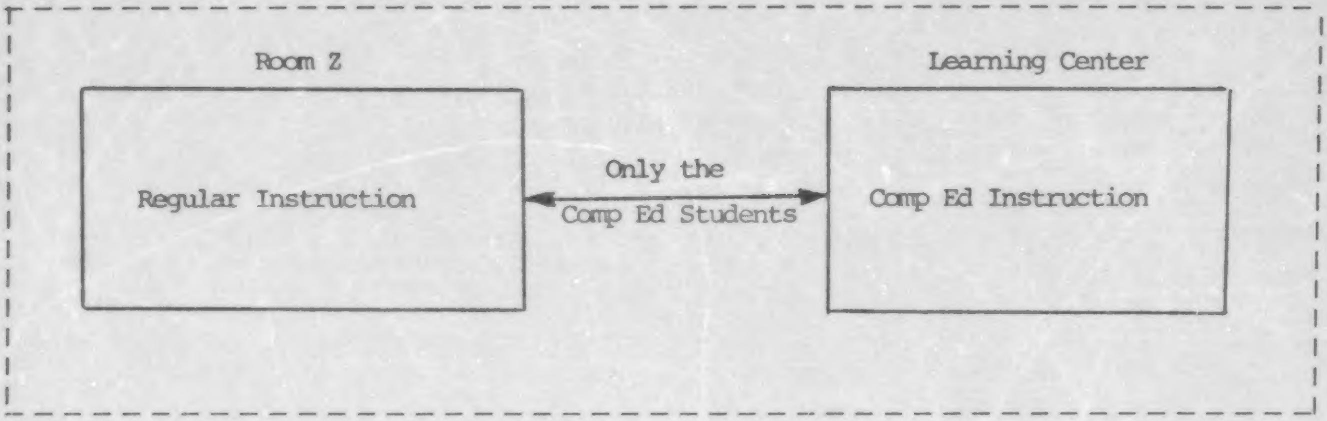
"PO"	Performance Objectives
"D/P"	Continuous Diagnosis/Prescription
"ALP"	Existence of Alternative Learning Paths
"PACE"	Individual Pacing

For definitional purposes, the matrix shown in Exhibit III-9 will form the basis for program classification. The cells of this matrix have been numbered from 1 through 16 as shown. The pattern of Yes/No responses to the four preceeding issues are also shown within each cell with the responses associated with that cell listed from top to bottom in the cell corresponding to PO on top, D/P next, ALP next, and PACE on the bottom.

Note that six of the 16 cells have been shaded. These are cells which contain two "Yes" plus two "No" responses. These shaded cells have

Exhibit III-8

EXAMPLE STUDY "CLASSROOM" FOR PULL-OUT PROGRAMS


Room Z
Regular Instruction
Only the
Comp Ed Students
Learning Center
Comp Ed Instruction

DEFINITIONAL MATRIX

		PO - YES		PO - NO	
		D/P - YES	D/P - NO	D/P - YES	D/P - NO
ALP - YES	PACE-YES	1 Y Y Y Y	2 Y N Y Y	3 N Y Y Y	4 N N Y Y
	PACE-NO	5 Y Y Y N	6 Y N Y N	7 N Y Y N	8 N N Y N
ALP - NO	PACE-YES	9 Y Y N Y	10 Y N N Y	11 N Y N Y	12 N N N Y
	PACE-NO	13 Y Y N N	14 Y N N N	15 N Y N N	16 N N N N

been excluded from the anticipated analysis since no clear pattern of individualization or standardization emerges. This is the case because the study design team chose to establish no hierarchy of importance among the four key issues reflected in the matrix. It should be clear that the design reflected in the overall matrix is factorial in nature. Indeed, if programs for all 16 cells could be located in a reasonably balanced manner, the factorial design reflected would allow a thorough evaluation of each of the four key issues (main effects) plus all the relevant interaction among these four key issues. If this were true, the analysis would be a study of these issues, and programs would not require further labeling. However, the study design team considered each of the cells in the definitional matrix with great care -- reflecting the wide program experience available within and to the team -- and determined that logically not all of the 16 cells are equally likely to be found (i.e., certain specific patterns of Yes/No responses to the four key issues would not exist too widely in practice). Thus, the full factorial power of the matrix was not to be available. Given this practical limitation, it was decided to establish the four study groups described below.

- Group I — all "Yes" responses; considered to be the most individualized programs given the structured definition of an individualized program as suggested here; includes only cell number 1.
- Group II — three "Yes" responses plus one "No" response; considered to be the next most individualized group of programs within the suggested definition; includes cells numbered 2, 3, 5, and 9.

- Group III one "Yes" response plus three "No" responses; considered to be almost totally standardized (or traditional or non-individualized) programs within the suggested definition; includes cells numbered 8, 12, 14, and 15.
- Group IV all "No" responses; considered to be the most standardized programs given the suggested definition, includes only cell number 16.

The four groups may be summarized as follows:

- Group I very individualized programs
- Group II individualized programs
- Group III standardized programs
- Group IV very standardized programs

The reader is cautioned from viewing this definitional matrix as one which somehow maps the distribution of actual comp ed practice. Such a view might lead one to conclude that excluding the shaded portion of Exhibit III-9 from the analysis would exclude a large number of programs from potential consideration. The study team does not believe this to be the case based on previous experience as well as specific experiences gained in identifying candidate sites during this design phase. However, even if it were the case, the RFP expressly states that NIE is not interested in a study of representative individualized practice in comp ed in the U.S. today. Such a survey -- useful for broad policy decisions -- is the subject of District Survey I. This study at hand is tasked to contrast well-implemented individualized programs and well-implemented standardized programs and the sample selected for study need not be reflective of the average experience in this regard in the U.S. today.

However, if the study implementation contrasts were to discover a significant number of potentially interesting programs falling into all 16 cells of Exhibit III-9, the reliance on the above defined four study groups could be reduced in favor of additional reliance on the factorial power of the basic matrix. The contrasts to be addressed would shift from comparisons of effectiveness between or among groups to a consideration of the main effects and interaction effects contained with the matrix obtained. The TURNKEY study design team does not feel that this is a likely event; it is mentioned here only in the interest of completeness.

Exhibit III-10 summarizes the design framework of the overall sampling structure.

Sampling Criteria

In addition to the framework into which all selected programs must fit, a number of specific criteria must be met by such programs. These criteria relate to the following considerations:

- Practicality -- the necessity for selecting programs for which data collection may be conducted in an efficient manner with LEA cooperation ensured;
- Design -- the requirement for programs which meet the design specifications outlined by NIE and/or included in the proposed design;
- Stability -- the need for selecting programs which reflect a settled program which is operating in a stable mode and has overcome start-up and disruptive problems;
- Implementation -- the requirement for available programs of sufficient scope that meaningful implementation data can be collected and verified prior to final selection.

EXHIBIT III-10

SPECIFIC DESIGN FRAMEWORK

A. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Mainstream
SUBJECT AREA: Reading
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I, II, III, Iv

B. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Pull-Out
SUBJECT AREA: Reading
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I, II, III, IV

C. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Mainstream
SUBJECT AREA: Math
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I, II, III, IV

D. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Pull-Out
SUBJECT AREA: Math
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I, II, III, IV

Exhibit III-11 describes the specific sample selection criteria which will be followed in the sampling process. It also specifies the general rationale(s) for the use of each criterion in the selection and identifies the relevant items in the two (district and school level) sampling questionaires which will be discussed later in this section.

Sample Selection Procedures

The procedures by which the final sample will be selected are dictated by the nature of the universe(s) from which the districts and schools are to be chosen.

According to NIE specifications for the study, the final sample is to contain programs selected from those included in NIE's District Survey I, as well as those chosen on the basis of the investigation during this study. In our interim progress report, submitted to NIE on 1 August 1975, we recommended that this specification be eliminated and that the sample selection procedures for the two studies be completely divorced. In the RFP and from conversations with NIE staff, it has been stated that a major purpose of this sample "overlap" between the two studies is to provide validation (in the form of classroom observations) of the data gathered in questionnaire form, during District Survey I, from teachers and administrators. Although spot validation of these questionnaire responses is clearly necessary, the advisability of using (and possibly confounding) the large-scale Individualized Instruction Study as a checking mechanism is open to question. It appears more scientifically sound for a validation mechanism to be built directly into the design of District Survey I or to

Exhibit III-11
Page 1 of 2

CRITERION	QUESTIONNAIRE ITEM(S) (D-DISTRICT) (S-SCHOOL)	RATIONALE			
		PRACTICALITY	DESIGN-REQUIRED	STABILITY	IMPLEMENTATION DATA
Willingness to participate must be indicated at both the LEA and the school level.	D-13 S-19	X			X
A school must either be a Title I school or a Title I eligible school.	S-7	X	X		
A program must have been in operation in the LEA for at least the 1974-75 and the 1975-76 school years, and in the school since at least the 1975-76 school year.	D-11(7), 12(7) S-14(7)			X	X
A program must have been continued from the 1975-76 school year to the 1976-77 school year in the study classrooms without major changes in operation.	D-11(8, 9) 12(8, 9) S-14(8, 9) 15(1, 2, 4)			X	X
A program must be in operation in at least 10 grade two classrooms and at least 10 grade three classrooms in the LEA and at both grade levels in the study schools.	D-11(2, 3) 12(2, 3) S-14(2, 3)	X		X	X
A classroom must contain a projection of at least 6 comp ed students in the program of interest for mainstream programs or at 6 comp ed students from any regular classroom served by the center for pull-out programs.	D-7, 8 S-8, 9 S-15(5)	X	X	X	
No recently implemented "major" student redistribution that has affected more than 15% of the enrollment in potential study schools (e.g., no busing orders or plans, no redistribution of students due to clsoing of other schools with declining enrollments) since 1 September 1974.	S-16			X	X

Exhibit III-11
Page 2 of 2

CRITERION	QUESTIONNAIRE ITEM(S) (D-DISTRICT) (S-SCHOOL)	RATIONALE			
		PRACTICALITY	DESIGN-REQUIRED	STABILITY	IMPLEMENTATION DATA
No classroom will be included in the study whose teacher is new to the program during the 1976-77 school year; more specifically, for mainstream programs, no classroom whose comp ed teacher is new to that classroom that year will be included and for pull-out programs none of the comp ed teachers serving grades 2-3 comp ed students can be new to the program that year.	S-15(4)			X	X
A school must have a non-minority population percentage which falls within 15% of the district's overall Title I school, non-minority population percentage.	S-13	X	X		
A program must have adequate documentation in the form of curriculum guides, suggested teacher lesson guides, or something similar and district or school-level mandates regarding these guidelines.	D-11(14) 12(14) S-14(14)	X			X
At least 80% of the study body in a school must speak English as their native language.	S-12	X	X		
No more than one generic type of comp ed program in reading and one in math per school building may be available in a study building.	S-14(1)		X	X	X
Both the district level and school level program administration must have checked at least the next to the highest level implementation/effectiveness response in their respective survey questionnaires.	D-11(6) 12(6) S-14(6)				X

commission a small validation survey as a separate entity. For a number of reasons, the requirement that 50% of the sample for the Individualized Instruction study be selected from the District Survey I sample may result in confounding of the results of both studies:

1) Because District Survey I data will be collected during the 1975-76 school year and the individualized instruction study data will be collected during the 1976-77 school year, it will be difficult to determine if apparent inconsistencies are really discrepancies or merely reflective of changing program characteristics (Note: our research indicates many individualized programs are undergoing continuous structural change).

2) Should validation by observation reveal apparent weaknesses in the District Survey I questionnaire procedure, no effective remediation process is available.

3) In the event that apparent discrepancies are uncovered, it is possible that unproductive disagreement between the two contractors regarding the efficacy of both questionnaire and validation techniques could occur.

4) It is possible that the District Survey I sample does not include enough sites which are suitable for inclusion in the Individualized Instruction study, in terms of program stability, size, or willingness to participate.

5) The validation role of the Individualized Instruction study may require that its data collection instruments be structured in a manner which is not in the best interests of the overall analysis plan.

6) Data availability from District Survey I may cause severe time constraints to be placed on the sample selection activities necessary for the Individualized Instruction study.

For these reasons, which were indicated in our interim report, we recommend that this sampling restriction be lifted from the study design. We hasten to note that this change would in no way preclude a District Survey I site from inclusion in the Individualized Instruction study, should it meet the appropriate selection criteria for both studies.

Whether or not the sampling ties between the two projects are severed, the sample selection procedures will remain essentially unchanged. Under both sampling designs, a preliminary data collection activity will be necessary, during the early part of project implementation, before final selection of programs can actually occur. For all programs included in District Survey I, relevant items will be included in the teacher questionnaire and possibly elsewhere for that study. For other potential candidate programs, preliminary questionnaires requesting the necessary information for the final selection of programs will be sent to appropriate school districts. Appendix 5 contains these suggested preliminary sampling questionnaires and their cover letter. Based upon the responses to these questionnaires, each potential program will be assessed relative to the sample selection criteria described earlier in this section. The thirteen selection criteria will be applied in the indicated priority order (as shown in Exhibits III-13 and III-14) to each potential program until the appropriate number of programs (including the required 25% oversampling) have been chosen to fill the appropriate cells of the sampling matrix. In addition, this preliminary sample will include an additional 25% oversample of programs, which will serve as replacements for selected programs found unsatisfactory upon on-site verification.

Exhibit III-12 depicts, in flow chart form, the sequence of sample selection activities. In this paragraph, we expand upon the activities noted in Exhibit III-12:

Exhibit III-12
Page 1 of 2

SAMPLE SELECTION PROCESS

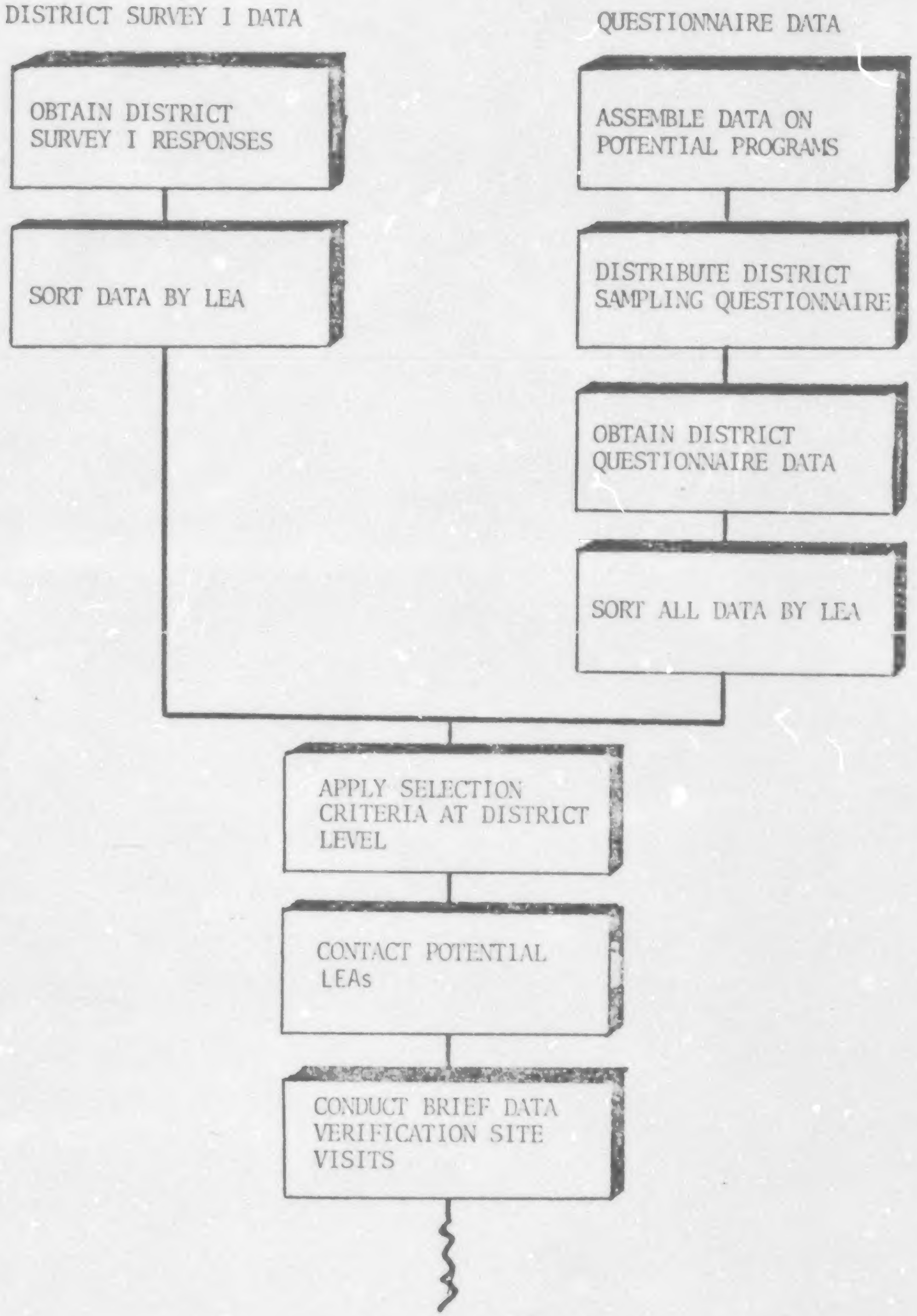

DISTRICT SURVEY I DATA
QUESTIONNAIRE DATA
OBTAIN DISTRICT SURVEY I RESPONSES
SORT DATA BY LEA
ASSEMBLE DATA ON POTENTIAL PROGRAMS
DISTRIBUTE DISTRICT SAMPLING QUESTIONNAIRE
OBTAIN DISTRICT QUESTIONNAIRE DATA
SORT ALL DATA BY LEA
APPLY SELECTION CRITERIA AT DISTRICT LEVEL
CONTACT POTENTIAL LEAs
CONDUCT BRIEF DATA VERIFICATION SITE VISITS

Exhibit III-12
Page 2 of 2

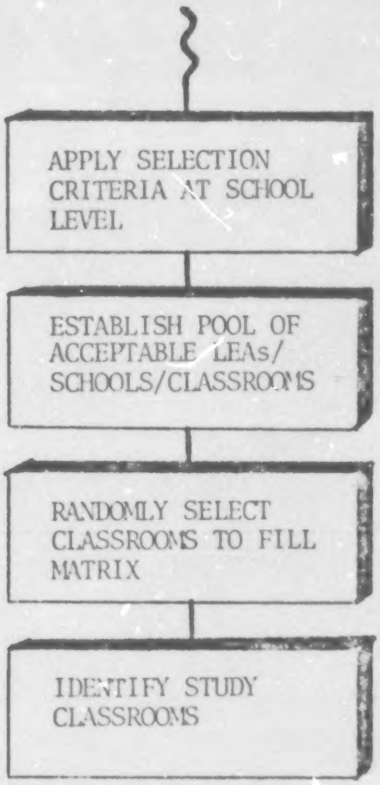
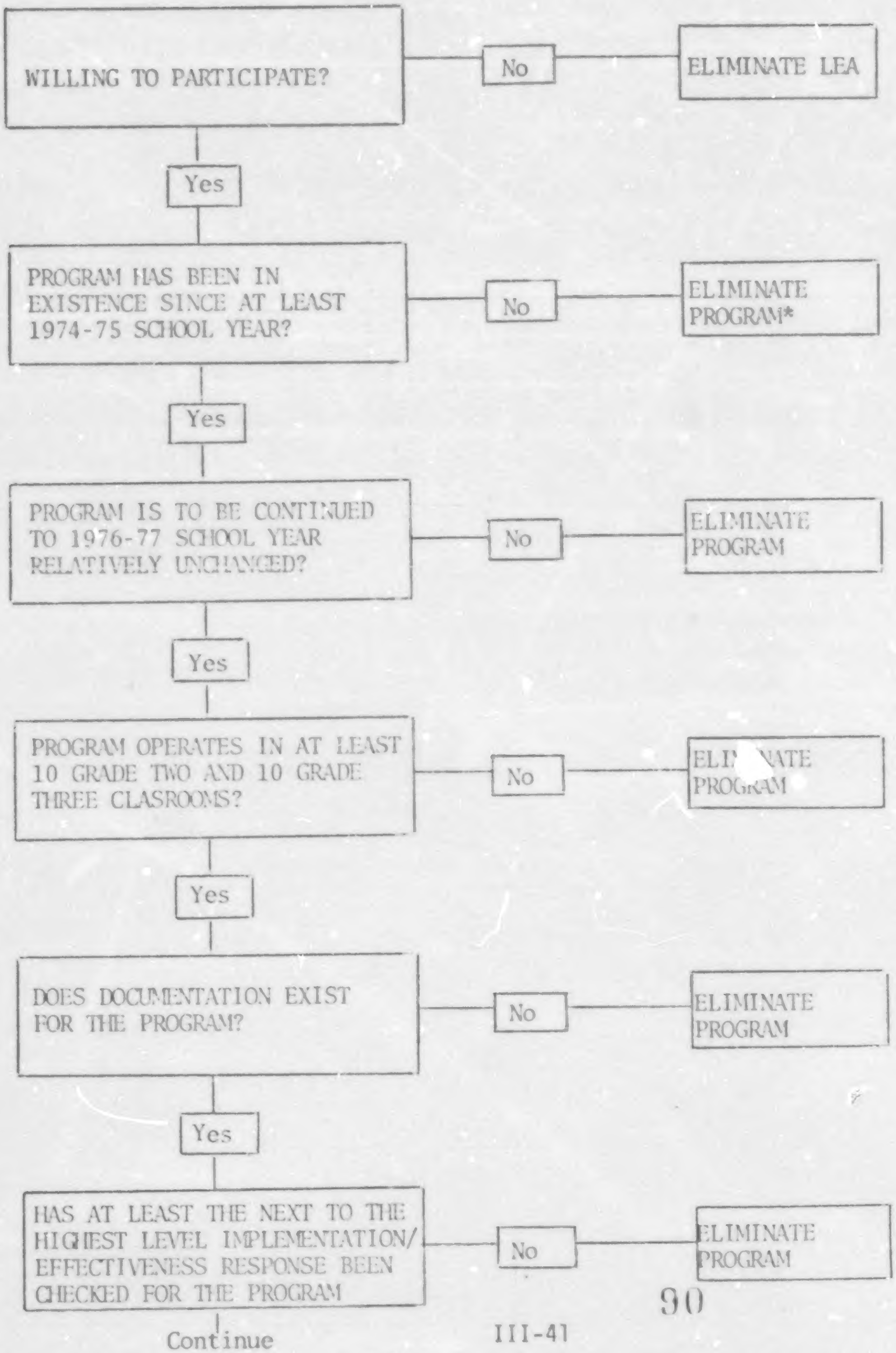

APPLY SELECTION CRITERIA AT SCHOOL LEVEL
ESTABLISH POOL OF ACCEPTABLE LEAs/ SCHOOLS/CLASSROOMS
RANDOMLY SELECT CLASSROOMS TO FILL MATRIX
IDENTIFY STUDY CLASSROOMS

Exhibit III-13

SELECTION CRITERIA APPLIED AT DISTRICT LEVEL


WILLING TO PARTICIPATE?
No
ELIMINATE LEA
Yes
PROGRAM HAS BEEN IN EXISTENCE SINCE AT LEAST 1974-75 SCHOOL YEAR?
No
ELIMINATE PROGRAM*
Yes
PROGRAM IS TO BE CONTINUED TO 1976-77 SCHOOL YEAR RELATIVELY UNCHANGED?
No
ELIMINATE PROGRAM
Yes
PROGRAM OPERATES IN AT LEAST 10 GRADE TWO AND 10 GRADE THREE CLASROOMS?
No
ELIMINATE PROGRAM
Yes
DOES DOCUMENTATION EXIST FOR THE PROGRAM?
No
ELIMINATE PROGRAM
Yes
HAS AT LEAST THE NEXT TO THE HIGHEST LEVEL IMPLEMENTATION/EFFECTIVENESS RESPONSE BEEN CHECKED FOR THE PROGRAM
No
ELIMINATE PROGRAM
Continue

*An LEA may have more than one program

Exhibit III-14
Page 1 of 2

SELECTION CRITERIA APPLIED AT SCHOOL LEVEL

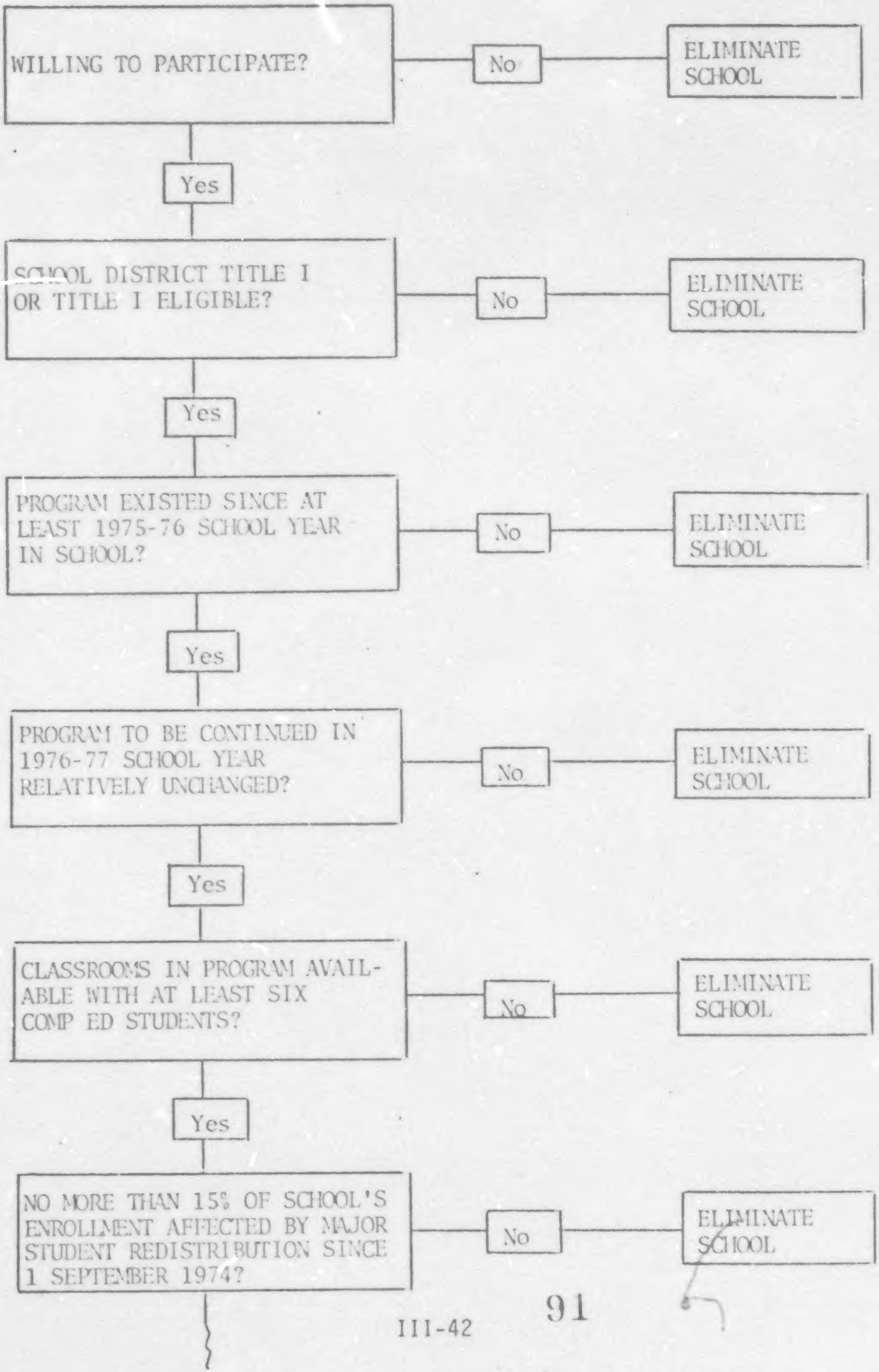
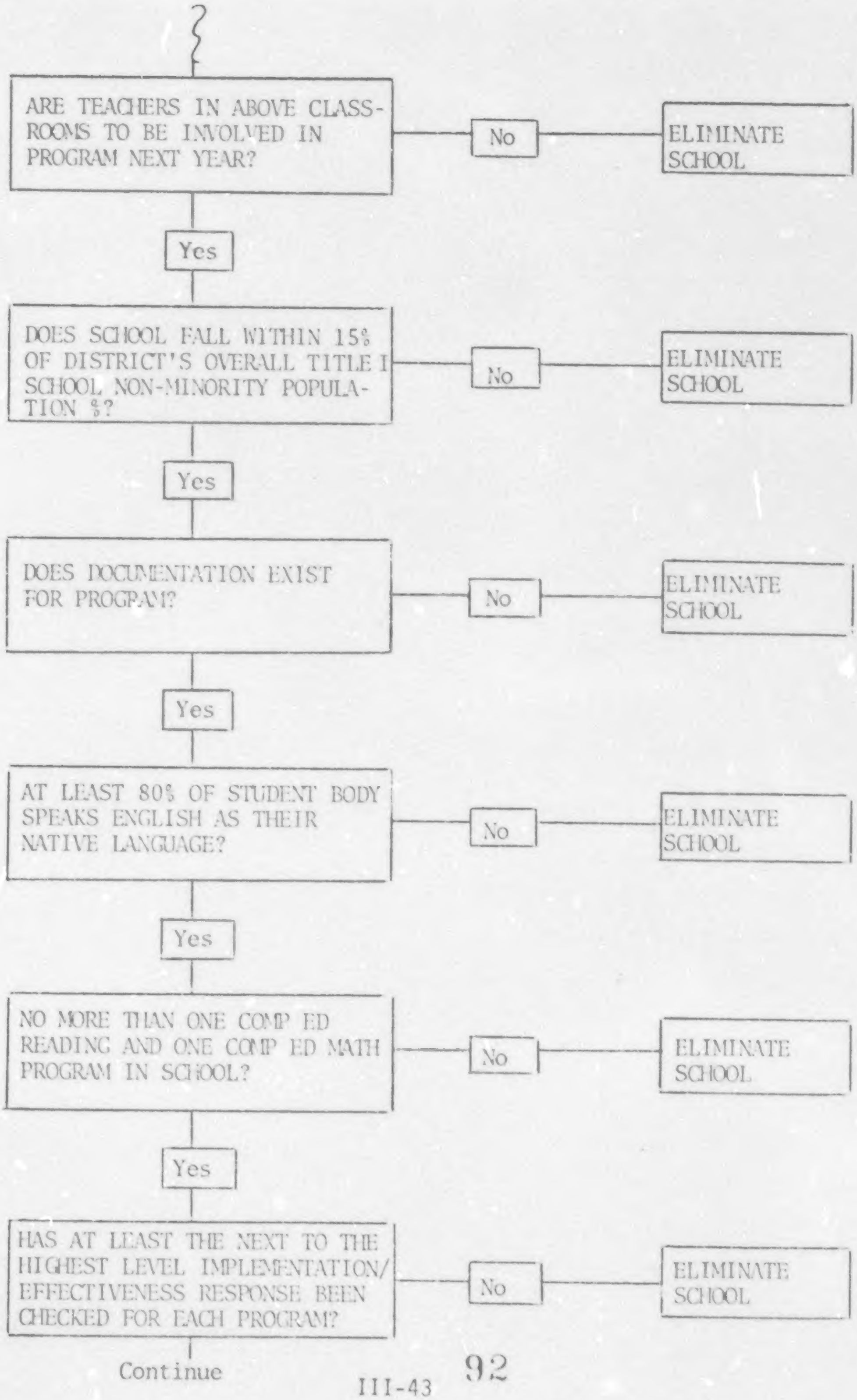

WILLING TO PARTICIPATE?
No
ELIMINATE SCHOOL
Yes
SCHOOL DISTRICT TITLE I OR TITLE I ELIGIBLE?
No
ELIMINATE SCHOOL
Yes
PROGRAM EXISTED SINCE AT LEAST 1975-76 SCHOOL YEAR IN SCHOOL?
No
ELIMINATE SCHOOL
Yes
PROGRAM TO BE CONTINUED IN 1976-77 SCHOOL YEAR RELATIVELY UNCHANGED?
No
ELIMINATE SCHOOL
Yes
CLASSROOMS IN PROGRAM AVAILABLE WITH AT LEAST SIX COMP ED STUDENTS?
No
ELIMINATE SCHOOL
Yes
NO MORE THAN 15% OF SCHOOL'S ENROLLMENT AFFECTED BY MAJOR STUDENT REDISTRIBUTION SINCE 1 SEPTEMBER 1974?
No
ELIMINATE SCHOOL

Exhibit III-14
Page 2 of 2


ARE TEACHERS IN ABOVE CLASSROOMS TO BE INVOLVED IN PROGRAM NEXT YEAR?
No
ELIMINATE SCHOOL
Yes
DOES SCHOOL FALL WITHIN 15% OF DISTRICT'S OVERALL TITLE I SCHOOL NON-MINORITY POPULATION %?
No
ELIMINATE SCHOOL
Yes
DOES DOCUMENTATION EXIST FOR PROGRAM?
No
ELIMINATE SCHOOL
Yes
AT LEAST 80% OF STUDENT BODY SPEAKS ENGLISH AS THEIR NATIVE LANGUAGE?
No
ELIMINATE SCHOOL
Yes
NO MORE THAN ONE COMP ED READING AND ONE COMP ED MATH PROGRAM IN SCHOOL?
No
ELIMINATE SCHOOL
Yes
HAS AT LEAST THE NEXT TO THE HIGHEST LEVEL IMPLEMENTATION/EFFECTIVENESS RESPONSE BEEN CHECKED FOR EACH PROGRAM?
No
ELIMINATE SCHOOL
Continue

- Obtain District Survey I Responses -- For all sites included in District Survey I, appropriate data will be obtained from SRI;

- Sort Data by LEA -- This District Survey I data will be sorted by LEA;

- Assemble Data on Potential Programs -- Based upon data accumulated through literature searches, printed matter, and materials provided by publishers, agencies, etc., all data on potential programs will be assembled;

- Distribute District Sample Questionnaire (see Appendix 5) -- The preliminary district-level sampling questionnaire will be sent to the superintendents of the appropriate LEAs;

- Obtain District Questionnaire Data -- Responses from the District sampling questionnaire will be received;

- Sort All Data by LEA -- All program data will be sorted by LEA;

- Apply Selection Criteria at District Level -- The thirteen sample selection criteria will be applied, where appropriate, to the data on LEAs. LEAs which do not meet these criteria will be eliminated. Potential schools within each acceptable LEA will be identified. Exhibit III-13 illustrates this procedure in further detail;

- Contact Selected LEAs -- The LEAs which are selected according to the selection criteria will be contacted to arrange for site visits;

- Conduct Data Verification Site Visits -- Each in which a selected program is located will be visited by a member of the project team;

- Apply Selection Criteria at School Level -- Appropriate selection criteria will be applied to the school-level data gathered during the site visit. Exhibit III-14 illustrates this procedure in further detail;

- Establish Pool of Acceptable LEAs/Schools/Classrooms -- A pool will be established of all classrooms (by LEA and school) which meet the sample selection criteria;

- Randomly Select Classrooms to Fill Matrix -- From the pool of acceptable classrooms, individual classrooms will be randomly selected, one at a time, until each working cell of the sampling framework is oversampled by 25%.

- Identify Study Classrooms -- From those classrooms in each cell, those which will be considered as primary study classrooms will be identified.

Final Sample Dimensions

Appendix 6 describes in detail the statistical process followed for determining the sample sizes presented in this portion of our report. To summarize this appendix, the following are the key assumptions that are reflected in the numbers presented here:

- The most demanding (from a statistical point of view) comparison that the study will be forced to make would involve a univariate t test between two independent groups selected from the 16 cells defined in Exhibit III-10; a significance level of .05 with a 2-tailed test has been assumed.
- Group means from our "classroom" units of analysis will be utilized. These means will be based on approximately 6 students for mainstream programs and approximately 10 students from pull-out programs.
- The effect size which is considered desirable to detect, i.e., educationally significant, is 0.5 standard deviation difference between two population means.
- A conservative estimate of the intraclass correlation of achievement scores (i.e., correlation of scores from students within the same classroom regardless of treatment) is 0.5.
- The level of statistical power (i.e., the probability of detecting a true difference of 0.5 standard deviation in two population means) used here is .90.

Based on these assumptions and the more detailed discussion contained in Appendix 6, Exhibit III-15 displays the suggested number of classrooms required for each of the 16 study cells implied in Exhibit III-10. The total number of classrooms included in this display is 800, with no allowance for oversampling. Using a 25% oversampling rate, a total of 1,000 classrooms would be included in our proposed study design.

EXHIBIT III-15

NUMBER OF CLASSROOMS NEEDED IN THE STUDY SAMPLE, WITH NO OVERSAMPLING*

	PROGRAM TYPE			
	GROUP I	GROUP II	GROUP III	GROUP IV
A. Mainstream Reading	51	51	51	51
B. Pull-Out Reading	49	49	49	49
C. Mainstream Math	51	51	51	51
D. Pull-Out Math	49	49	49	49

*Using a 25% oversampling factor, a total of 1,000 classrooms are needed.

Recalling that study Groups II and III each include four cells from within our definitional matrix while study Groups I and IV each include only one cell for this matrix, it is apparent that not each individual cell in our definitional matrix need be equally represented to fulfill the sample dimensions listed here. However, to the extent possible, the individual cells within each of the multi-celled study group should be represented in a relatively balanced manner to facilitate any contrasts to be made which call for data to be disaggregated to that level of detail (see Task 5, Level 2 Analyses). For the overall study design constraint, however, it is only important that the numbers listed in Exhibit III-15 be adhered to, if the sample size determiniation process used here is to be relevant.

The earlier references to proportional sampling should be considered at this time. If the implementation contrast should find that for instance overall only 30% of the programs available for this study are of the pull-out variety or that only 10% of the available programs can be classified as Group IV (neither of these illustrations are felt to be accurate, however), the number of required sample classrooms reflecting these dimensions could then be adjusted to reflect these reality limits. This adjustment could be made by maintaining all those feasible or readily attainable treatment categories at the level shown in Exhibit III-15 while adjusting downward the sample requirements not-so-readily attainable treatment categories such that the ratio of the non-adjusted sample sizes to the adjusted ones reflects the ratio the numbers of such programs overall. In the first of the hypothetical adjustments indicated by our illustration, the second and fourth lines of Exhibit III-15 would be reduced from 49 to 21 to reflect the 70/30 ratio indicated in program availability data. In the second such adjustment, the fourth column of Exhibit III-15 would be changed from 51, 49, 51, 49 to read 17, 16, 17, 16 to reflect the 30/10 ratio between each of the other three columns and the fourth column indicated in the program availability data.

No need for reliance on proportional sampling as described above has been identified to date. The difference in sample size between the mainstream and pull-out programs reflected in Exhibit III-15 is solely based on the differential assumptions used for the number of students from each type of classroom setting to be tested for purposes of this study, not on any assumption or data on the basic availability of programs from either of these treatment strata.

Additionally, it should be recalled that the study design calls for approximately the same number of nominally second grade classrooms as nominally third grade classrooms reflected in each treatment category. This means that about 24-26 classrooms at each of these grade levels would be included in each program grouping.

As a closing note on the suggested sample dimensions, it may be estimated that the 1,000 classrooms discussed in this section will reflect the following numbers of programs and LEAs:

- Assuming 10 classrooms per program on the average, a total of approximately 100 programs will be studied.
- Assuming 4 programs per LEA, on the average, a total of approximately 25 LEAs will be involved in the study.

***

The nature of the sampling plan, especially its reflection of our definitional matrix, allows a number of issues critical to NIE and the prime audience of the proposed study (i.e., Congress) to be addressed in a highly structured manner. Though a true experiment (wherein all else is kept constant while only one program feature is systematically varied at any given time and students are assigned to treatments randomly) is impossible to achieve in any real world setting, the quasi-factorial structure suggested here allows an approximation of at least a portion of this scheme to obtain. (The reason for not fully utilizing the factorial-interact potential of our definitional matrix was stated earlier.) Careful selection of programs and matching of students in adherence to the sampling contrasts listed earlier will suffice as the suggested approximation of random assignment of students, recognizing that this approximation has been a major problem for all large scale field research and evaluation studies in education.

TASK 3 - DEVELOP PROCEDURES FOR ASSESSING PROGRAM IMPLEMENTATION

Summary of Critical Issues

In our Proposal, we identified (and discussed rather extensively) a number of critical issues regarding the assessment of program implementation. In addition, we discussed several proposed approaches to resolving these issues during the design phase.

First, the critical issues identified and discussed in our Proposal included:

- The need conceptually to separate well-implemented programs from highly-individualized ones (i.e., a standardized program could be well-implemented) from highly effective programs, recognizing that interaction exists between these measures (i.e., a well-implemented plan may not be a highly individualized program, and conversely, a highly-individualized program may not be well-implemented.
- The difficulty inherent in obtaining/developing a baseline regarding the most effective and valid prescribed plan for implementation, in light of contextual restraints (e.g., LEA budget constraints precluded the purchase of all supplemental materials designed for use in an individualized reading center), the intentions of publishers (e.g., the desire to accommodate teacher desires as a trade-off against effective implementation), and other factors.
- The need to resolve an apparent conflict between the inclusion of only well-implemented programs in the study and then assessing the impact of varying degrees of implementation on program outcomes (see RFP).

In addition, during the design phase, we identified several additional critical issues, including: a) whether or not a single uniform instrument could be used to determine in a valid manner the degree to which specific individualized programs are well-implemented; and b) the degree to which cross-program comparisons relating degree of implementation to outcomes could be made.

Second, implementation, in the context of the RFP, was taken to mean some combination of stability over time (i.e., steady state operations beyond the transiency of any start-up activities), plus operation in accordance with the plan as initially set out or as modified prior to the school year under study, 1976-77. This view of implementation lends itself to a two-step process for ensuring that only those programs that are likely to have been well-implemented are selected for study. Subsequent verification through on-site observations and other data collection procedures will determine whether the programs studied are indeed well-implemented, and if not, the reasons why.

As described in the Proposal, a number of sub-steps were discussed in attempting to resolve the above issues as they relate to initial identification, including review of existing program implementation guides, results of studies with similar purposes (e.g., the SRI Follow Through study), and initial discussions with LEA staff members involved in candidate projects. A great deal of time was expended in attempting to resolve issues surrounding the second step (i.e., determining the degree to which potentially well-implemented programs are indeed well-implemented). In the following pages we discuss the activities undertaken during this task, the relevant findings and results of our efforts, and the procedures used in developing instruments used during the design phase and those for developing additional instruments to be used during the implementation phase of the study.

Approach Taken/Findings

During the design phase, procedures and instruments (checklists) were developed to identify potential candidates and determine tentatively whether or not candidate programs were well-implemented. Instruments were also developed to verify whether or not potential candidates met minimal criteria for inclusion in the study (see Appendix 5). Lastly, instruments and procedures must be fully developed and field tested for collecting data (through observation, surveys, and audits) on the implementation of programs during the school year, 1976-77. Under this task we discuss the procedures and instruments which have been developed and used thus far in the design phase as well as the specifications and procedures for developing additional instruments during the implementation phase. It should be noted that the overlap between this task and the previous task, sample selection, and with the site selection procedure described in Appendix 7 is extensive: to eliminate redundancy, appropriate references have been made for the reader.

Identification of Well-Implemented Candidate Programs

As described in Appendix 7, upon design contract award the TURNKEY project staff initiated several procedures to identify potential candidates whose programs, individualized or standardized, were well-implemented. Since obtaining OMB clearance for a formal survey during the design phase was considered infeasible, existing documentation and knowledge had to be heavily relied upon. Minimal criteria for inclusion included: a) that the program covered grades K-4, math and/or reading, and focused on compensatory education target groups; b) that documentation was readily accessible in the form of program descriptions, evaluation reports, validation findings, panel reviews, or personal observations of the project team members. A list of potential candidates (see Appendix 8) was gathered from: Right To Read demonstration program (USOE); USOE Title I and Title III offices, including programs which were or are presently being reviewed by the Dissemination Review Panel (DRP) of USOE; sites and potential sites for inclusion in the Program Information Package program (USOE); and the TURNKEY documentation file of almost 100 projects in 20 states. In addition, the major publishers of individualized learning systems were contacted and their nomination of the five districts which were implementing their program in the most effective manner was solicited.

In order to determine whether potential candidates were "likely" candidates, meeting minimal criteria, checklists were developed. The first, enclosed in Appendix 9, was developed in August and submitted to NIE for inclusion in District Survey I as an instrument to assist in sample selection. This instrument was similar in most respects to a second instrument developed by TURNKEY staff and used to determine whether potential (non-Survey I) sites met minimal criteria and thus qualified as likely candidates. This instrument went through two revisions and was reviewed by three LEA consultants and other education officials, to determine its practicality and feasibility in light of the nature of generally existing documentation. This checklist is enclosed in Appendix 10.

Next, the checklist was applied to potential candidates by members of the project team. In all instances, at least half of the information needed to complete the checklist had to exist in documentation available to project team members. Upon partial or full completion of the checklist, project team members talked with individuals responsible for or knowledgeable about the respective programs at the respective sites to complete the checklist and/or to verify certain portions of it. Upon completion of the checklist procedures the likely candidates were assigned to the respective sample cells, described in Task 2. In addition, the information enclosed in the identification checklist (Appendix 10) was punched on individual Keysort cards as likely candidates for future consideration.

We are aware of some of the limitations of the above procedures for identifying well-implemented programs. However, in light of the time and other constraints mentioned earlier, this procedure was felt to be the most feasible. First, we assumed that the publishers of programs who have prescribed sets of implementation guides and/or checklist procedures have designed them to insure greater program effectiveness and/or acceptance by teaching staff. While we feel comfortable with the sites nominated by the publishers, we are also aware of the need to review implementation plans to ensure that "suggested" procedures do increase the probability of high student performance rather than increased sales.

Second, Title I and Title III projects which have been validated and, particularly, those which have been approved by the Dissemination Review Panel, have received rather extensive evaluations over at least two years and in most instances have been validated not only by federal officials but also by state education officials. A review of validation forms and procedures used by USOE and several SEAs indicates a strict adherence to documented plans and objectives and the degree to which the programs followed plans and met objectives to be a relatively good baseline for assessing implementation. As described later, however, it is important to note that verification through survey and on-site observation prior to final selection will be necessary, as described in Task 2 and below.

In an attempt to develop procedures and instruments for assessing the degree of program implementation during the study implementation phase, we undertook the following tasks: a) a review of "critiques" of approximately 20 commonly-used math and reading systems which could be or were predominately implemented in an individualized mode (EPIE Reports 46 and 65); b) a preliminary analysis of the implementation guides/manuals provided by publishers, regarding specific learning systems; c) an analysis of observation and other checklists used by publishers' consultants and implementation checklists developed by external groups (e.g., /I/D/E/A/ Kettering regarding the IGE program); and d) a review of recent studies which have attempted to identify implementation variables and assess the degree to which implementation differed among the various programs, particularly the SRI Follow Through study.

In addition, project members had an opportunity to discuss the problems and approaches in assessing implementation with: a) three representatives from /I/D/E/A/who shared their time, internal checklists, and the unpublished results of their study,utilizing these checklists with the project members; b) Federal officials and SRI staff involved in the Follow Through study; c) numerous Follow Through directors whose models were assessed on the SRI implementation scale; d) Federal and other consultants who provided technical assistance during the implementation phases of the Right To Read program; e) LEA consultants, including principals and other staff, involved in the implementation of Title I programs; and f) several consultants responsible for training Title I teachers and other staff in the implementation of a variety of individualized learning systems.

Based upon these activities and reviews, the advantages and disadvantages of two approaches for assessing program implementation during the operational phase of the project were considered: a) whether or not it is practical, feasible, and justifiable to develop one instrument which could be used to assess degrees of implementation across all programs, standardized and individualized, and across all sites; or b) whether a unique instrument and set of procedures should be developed for each of the programs in order to assess in a reliable and valid way the degree to which a specific program was being implemented across individual classrooms utilizing that program. The implications for data collection and analysis of the two alternatives are rather significant.

First, if one could develop a uniform instrument for assessing implementation across programs and sites, and if such an instrument would provide reliable and valid data for use in sample selection, verification, and data collection during the period of observation, then the data collected could be used in the analysis as an independent variable which could possibly be associated with several outcome or dependent variables across all programs.

Second, if one were to develop unique instruments to assess implementation for each program, the lack of a common instrument would erode somewhat the validity of assessing degree of implementation across programs. However, information gathered through surveys, program audit checklists, and observation, based upon prescribed implementation guides, would probably be more reliable and valid with respect to the degree to which a specific program was being implemented according to a predetermined plan. While the instrument in the former alternative would have to be global in nature, the procedures developed under this alternative would be less obtrusive (i.e., through the review of documented evidence such as student progress check records); observations could be reduced, thereby avoiding anxieties on the part of teaching staff, and would probably be less time-consuming than the observations which occurred during the SRI study.

Both of these alternatives have several inherent problems. First, since the final program selection will not occur until the spring of 1976, no general instruments for the selected programs could be done with certainty. (However, as a surrogate, one could select programs from potential candidates and assess their respective implementation procedures.) Second, discussions with Follow Through staff at the respective Follow Through model development sites indicate that, with the exception of the LRDC model, most programs are undergoing revision at the present time. Hence, revisions of the instruments would probably have to be made during the implementation phase of this study. Third, dependence solely upon implementation guides supplied by publishers rather than reliance upon actual implementation plans as modified by LEA staff is rather tenuous. Based upon project team knowledge and experience in a large number of LEAs where individualized learning are presently being used, the adaptation process has frequently resulted in major changes in procedures. For example, in Grand Rapids, token economies are used extensively in programs where they are not necessarily suggested or prescribed. In Dallas, instructional programs have been modified to accommodate the district-wide criteria-referenced testing system. In short, while publisher implementation guides provide a starting point, verification will require discussions with LEA staff, at which time the possibility of revision of instruments must be assessed. In order to select either of the above alternatives or some suitable combination, project staff members spent an extensive amount of time reviewing existing instruments and studies and conducting discussions with the individuals noted above in Appendix 8. From these discussions the following findings and observations appear to be relevant.

Findings

1. Existing instruments used to assess degrees of implementation are inadequate in terms of comprehensiveness, appropriateness, and degree of reliability, and validity, and adaptability for the purposes of this study.

The specific instruments and procedures reviewed for potential application adaptation included: the /I/D/E/A/ IGE implementation kit: the SRI Follow Through instruments; and 10 individual implementation guides and checklists recommended or used by publishers' consultants and/or LEA staff in assessing the degree of implementation of their individualized learning systems.

IGE Implementation Kit

Over the last few years, /I/D/E/A/ Kettering staff assigned to monitor the implementation of IGE throughout the country have developed an implementation guide or kit which specifies 35 outcomes that become the central focus of continuing in-service and implementation. The instrument is rather comprehensive in assessing process objectives which should occur during the planning and implementation of IGE programs over a period of time. Based on discussions with /I/D/E/A/ officials and a review of evaluations sponsored by /I/D/E/A/, it would appear that the items are appropriate for the IGE program. It should be noted that IGE is generally a large network, consisting of a "league" of schools which agree to certain conditions prior to adoption and implementing IGE. Also, the IGE process focuses upon rather global processes related to school organization, involvement of parents and the community, and staff development in a variety of individualized activities. The implementation checklist does not focus upon the implementation procedures of a specific program: rather, in many cases, an individualized instruction program such as IPI, with its prescribed procedures, will be implemented in an IGE school. In certain instances, conflicts exist between the IGE processes and those suggested in structured individualized packages. The IGE implementation kit provides for a number of processes which appear to be appropriate for assessing implementation in this study. However, for the most part, the instrument is limited to surveys rather than observation, does not focus specifically upon the classroom implementation procedures, and is structured around the IGE approach.

SRI Follow Through Instruments

The recent SRI Follow Through study represents the most intensive attempt to assess program implementation that has ever been conducted, although "differences in implementation" was only one of the concerns addressed. The survey and observation instruments, developed through a rigorous process, were designed to answer a number of other questions in addition to program implementation. Once these variables, including the implementation variables, were developed, an attempt was made to identify (through consultation with the model developers) the specific implementation variables which appeared to be critical to effective implementation of the respective Follow Through models. The unit of comparison was based upon the implementation procedures used in a "standard classroom." Hence, rather than determining how well each Follow Through model site implemented the Follow Through model program in accordance with a pre-determined plan, the study attempted to determine what differences, if any, existed between implementation procedures of Follow Through when compared to a standard classroom program. In addition to extensive classroom observation, a teacher survey instrument was also used with Follow Through participants. Resulting data, however, was found to be inadequate for analysis purposes.

As noted earlier, the Follow Through instruments (mostly observational) were not comprehensive enough, although some attempts to gather additional information were made. For example, parental involvement, as part of the Follow Through requirements, was observed only if parents performed services as a participant in classroom activities. IN cases where "at-home" instruction was provided in a number of the Follow Through models, observations were not conducted. Perhaps most importantly, while attempts were made, data was not available to determine whether differences in project preparation activities prior to the implementation of the programs occurred (e.g., planning, training). In the Michigan Cost-Effectiveness study, TURNKEY found that a number of variables such as the decision-making power over selection of materials delegated to teachers, the nature of teacher training (e.g., in diagnosis and developing performance objectives), the extent of training prior to program implementation, and other planning activities discriminated significantly between successful and unsuccessful programs. Any instruments designed to determine the degree of implementation must also take into account the pre-implementation activities which have occurred, the qualitative nature of these activities, and to the extent possible, the relative impact on effective implementation.

Even through the procedures used in developing the Follow Through instruments were extremely rigorous and were determined to be reliable through extensive training and inter-rater reliability checks, a serious question has been raised by a number of individuals involved in developing and providing technical assistance to Follow Through sites about the validity of these instruments. In addition, some of the unexpected findings (e.g., the negative correlation between praise and student performance) also raise questions regarding the validity of data collected through observations. One Follow Through model developer indicated that several errors of omission occurred in the sense that the number of implementation variables provided to him for selection and prioritization (e.g., critical vs. important) did not reflect certain activities which were uniquely critical to the model. In other instances the variables from which the selection had been made did not adequately describe the specific manner in which the activity was to have been conducted. In at least five instances, LEA or model development directors who were involved in the SRI Study did not want to be considered as a candidate site for this study due to their "bad experiences" with the Follow Through study. In light of the great care taken by SRI, the extent that the Follow Through instrument was lacking in validity (i.e., it did not assess what was supposed to have occurred) accentuates the difficulty of developing a single instrument that can be applied uniformly across programs which have different implementation plans.

Selected Specific Implementation Guides/Checklists

At this writing the project team has reviewed ten pre-implementation guides (e.g., teacher guides, supervisor guides, and administrator guides) which are included either in a pre-service training program or as part of the overall information package for an individualized learning system. The specific projects whose guides have been analyzed in a preliminary manner include:

a) Project TEEM, University of Arizona Follow Through Model; b) New Century (formerly Appleton, Century, Croft) Math and Reading System; c) the five program implementation components of the five Program Information Packages (PIPs) presently being field tested by USOE; d) High Intensity Learning System (Cohen/Random House System); and e) SRA Individualized Math System.

Feasibility Testing of Instrument Design

Based upon a review of these program guides, the project team developed a list of general categories of implementation variables. In addition, the various types of data collection instruments feasible for collecting data on the specific categories of implementation variables was also identified. These variables are presented in Exhibit III-15 which also identifies the type (e.g., interview) and purpose (verify vs. initial collection) of data collection instruments which could be used.

Upon identification of these variables we designed a preliminary observation instrument, following a format similar to that used by SRI in the Follow Through study and by Medley in devising the PROSE and OSCAR instruments. Basically this instrument identified "who" (e.g., teacher, student, teacher); "how" (e.g., verbal, written instructions, etc.); and "moderators" (e.g., praise, dispraise, neutrality) as indicated in Exhibit III-16. The answers to such questions were designed to assess events and to describe observed interactions and classroom activities. We were concerned, however, that such an instrument would lack appropriate validity and discrimination qualities. We, therefore, conducted a number of simulations and limited classroom observations to determine

Exhibit III-15
Page 1 of 2

CATEGORIES OF ACTIVITIES/EVENTS	TYPES OF INSTRUMENTS WHICH COULD BE USED		
	INTERVIEW FORM	OBSERVATION	CHECKLIST
I. PROJECT PREPARATION			
A. Project Planning			
1) Parent Involvement	X		X
2) Program Goals	X		X
3) Program Design:			
a) Select Commercial Program	X		X
b) Developmental	X		X
c) Combination of a) and b)	X		X
4) Establish Responsibility/Authority	X		X
5) Identify/Select/Procure Resources			
a) Staff	X		X
b) Facilities	X		X
c) Special Equipment/Materials	X		X
B. Staff Training			
1) Development of Training Program	X		X
2) Involvement of Students/Staff/Parents	X		
3) Logistics			
a) Scheduling	X		X
b) Time Allowed	X		X
c) Resources Needed	X		X
4) Nature of Training Sessions	X		X
II. PROJECT OPERATIONS			
A. Diagnosis/Testing	X	X	X
B. Prescribing/Assigning	X	X	X
C. Instructional Planning			
1) Role of Individual or Group Instruction within the System	X	X	X

Exhibit III-15
Page 2 of 2

CATEGORIES OF ACTIVITIES/EVENTS	TYPES OF INSTRUMENTS WHICH COULD BE USED		
	INTERVIEW FORM	OBSERVATION	CHECKLIST
II. PROJECT OPERATIONS (Cont'd)			
C. Instructional Planning (Cont'd)			
2) Media of instructional planning	X	X	X
D. Instructional Management			
1) Student Records/Information Retrieval		X	X
2) Information Sources Used for Instructional Decisions	X	X	X
3) Nature/Frequency of Decisions	X	X	X
E. Classroom Management			
1) Use of Resources	X	X	X
2) Organization of Learning Environment		X	X
3) Student Movement in Learning Environment		X	X
F. Student Motivation	X	X	
III. PROJECT EVALUATION			
A. Involvement of Students/Staff/Parents in Evaluation	X		X
B. Types of Evaluation Recommended for Project	X		X
C. Data Collection	X		X
D. Analysis	X		X
E. Reporting	X		X
F. Use of Evaluation	X	X	X
1) Feedback into the System	X	X	X
2) Other	X	X	X

Exhibit III-16

Observation Design

WHO

T	Teacher
A	Aide
V	Volunteer
S	Student
G	Small Group
L	Large Group
P	Principal
R	Reading Specialist
C/R	Computer Readout

WHAT (ACTIVITY)

D	Diagnosing
A/P	Assigning/Prescribing
Q	Questionning
A	Answering
L	Listening
O	Observing
R	Recording

TO WHOM

T	Teacher
A	Aide
V	Volunteer
S	Student
G	Small Group
L	Large Group
P	Principal
R	Reading Specialist
C/R	Computer Readout

HOW

V	Verbal
W	Written
D	Drama
P	Pagentry (Physical non-language)

MODERATORS

P	Positive
N	Negative
O	Neutral

whether or not the recorded responses would adequately assess the qualitative nature of "degree of implementation," recording in narrative form either the actual practice observed in the limited classrooms or suggested procedures described in implementation guides. Exhibit III-17 describes the result of these "field test" activities.

Suggested and Optional Procedures

In light of the limitations of the above approach, we have refined the second alternative -- the use of a specific instruments and implementation criteria for assessing each learning system used -- and designed some preliminary instruments to illustrate the approach. In addition, we developed an optional procedure and instrument, the feasibility of which will have to be tested during the implementation phase, to assess among programs the degrees of implementation only along the dimensions of individualization specified in the sampling design (e.g., the 16-cell matrix). A major difference between the suggested procedure and the optional procedure is the degree of involvement of the program developer in determining the instrument which is to evalute degree of implementation along with specific dimensions. With the suggested procedure, the developer has the major responsibility for defining the questions and observations within categories specified by the contractor. If the optional procedure is used, the developer is more limited in determining questions and observations, but still must define the range of acceptable answers and verification procedures which could be unique to his program -- a procedure similar to that used by SRI in the Follow Through study.

Exhibit III-17
Page 1 of 7

QUASI FIELD TEST OF OBSERVATION INSTRUMENT

The following is an example of the problems which occurred when trying to use a uniform observation checklist on five different existing learning systems which were in operation in potential target sites. The five programs considered have all met the eligibility criteria discussed in Task 2. These programs are listed below according to the following notation:

ALPHA	An Alpha II program in Grand Rapids, Michigan
LU	A Learning Unlimited Program in Grand Rapids, Michigan
PLAN	A Westinghouse PLAN program in Grand Rapids, Michigan
GINN	A Ginn 360 program in Grand Rapids, Michigan
M.C.	A Middle Cities program in Lansing, Michigan

I.

An area of the activities/events listed in Exhibit III-16 which were basic to all individualized programs was selected.

II. Project Operation

A. Diagnosis/Testing

The activities were recorded on an observation form.

Activity 1: Student Takes Diagnosis Test

PROGRAM	WHO	WHAT	TO WHOM	HOW	MODERATION
ALPHA	T	D	S	W	O
LU	T	D	S	W	O
PLAN	T	D	S	W	O
GINN	T	D	S	W	O
M.C.	T	D	S	W	O

Activity 2: Results of Diagnostic Test are Recorded

PROGRAM	WHO	WHAT	TO WHOM	HOW	MODERATION
ALPHA	A	R	S	W	O
LU	A	R	S	W	O
PLAN	A	R	S	W	O
GINN	T	R	S	W	O
M.C.	A	R	S	W	O

Exhibit III-17
Page 2 of 7

Activity 3: Student Receives Prescription Based on Diagnostic Test

PROGRAM	WHO	WHAT	TO WHOM	HOW	MODERATION
ALPHA	T	A/P	S	W	Nu
LU	C/P	A/P	S	W	Nu
PLAN	C/P	A/P	S	W	Nu
GINN	T	A/P	S	V	Nu
M.C.	T	A/P	S	W	Nu

Using this type of observation format, little difference was noted between the various programs.

II.

By interacting with site staff and program developers and reviewing implementation guides, it was determined that major differences did, in fact, occur in this simple diagnostic-prescription step. These differences would be important to note between programs, but more critical they must be noted by those responsible for ensuring that effective implementation of a program has occurred.

The following series of diagrams depict significant differences in the observed programs which were not identified by a simple cross program observation checklist.

SYMBOLS FOR EXAMPLE

A = Assigned Material

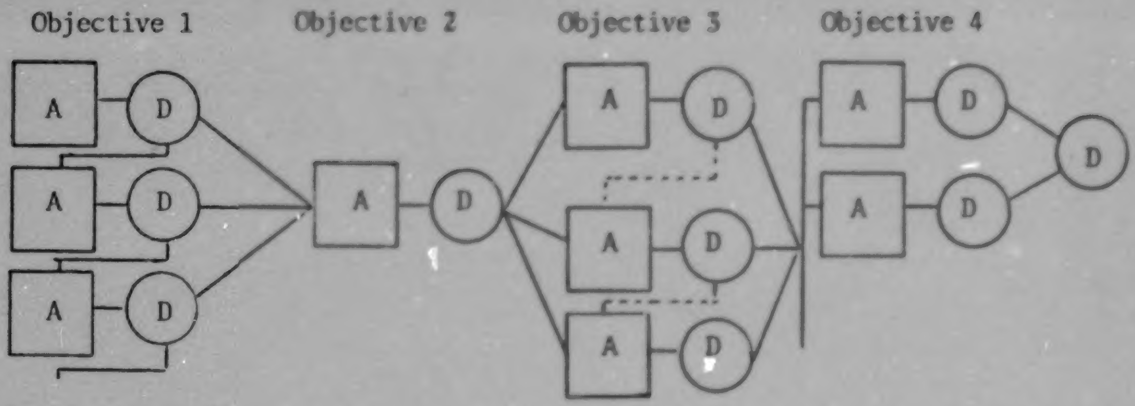
✓ = Test which is not used for Diagnosis (changing the students assignment) but only to ensure he is accurately responding to the material

T = Test which measures an objective but does not change assginment

D = Test which changes an assignment (i.e., Diagnosis for Prescription)

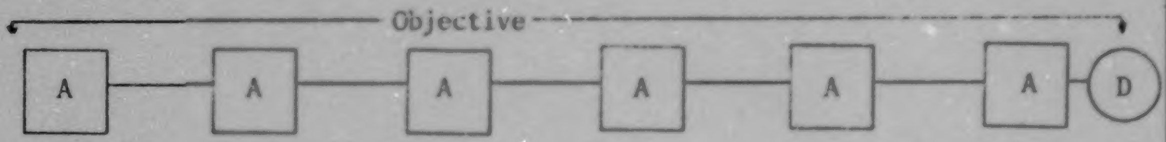
Exhibit III-17
Page 3 of 7

ALPHA Block approximates a 2-week work assignment


Objective 1
Objective 2
Objective 3
Objective 4
A
D
A
D
A
D
A
D
A
D
A
D
A
D
A
D
A
D
D

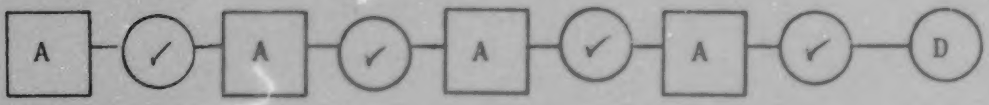
Small objective units - inter-modual branching possible.

PLAN TLU approximates a 2 week assignment


Objective
A
A
A
A
A
A
D

No small objective units. Branching to D test only if student "feels" he wants the "challenge".

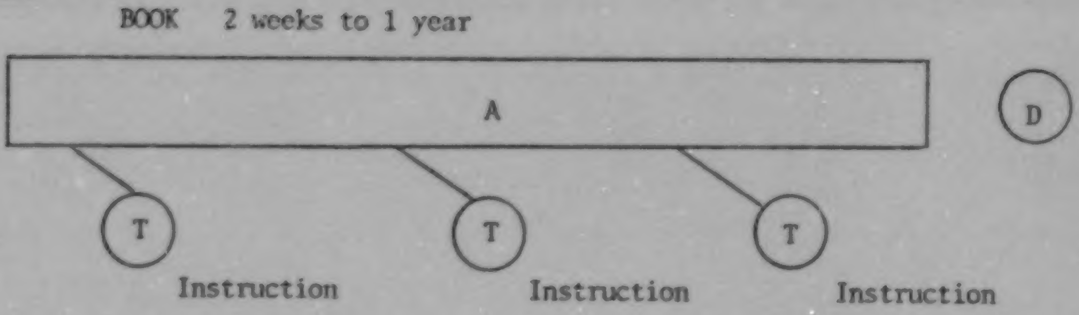
LU Module approximates a 2 week assignment



Small check tests - not used for branching due to "spin off" learning.

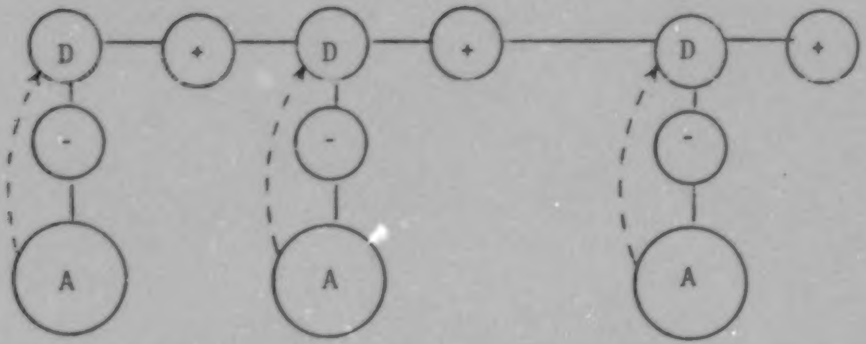
Exhibit III-17
Page 4 of 7

GINN



Reading assignment does not change after skill tests - only recommended instruction by teacher or assignment of supplementary ditto's. Assignment Book not consistant with skills tested.

M.C.



No sequence of tests required. Primarily evaluation with "available" cross referenced materials. All instruction noted accomplished by paraprofessionals, outside of regular class environment.

The major differences between programs can be specified only when the frequency of the diagnostic testing is determined and the consequences of the testing. In one system (LU) daily tests are used to ensure that the student is responding to the system. With good classroom control some site personnel feel they could be omitted and the system could be considered well-implemented. In a second system (ALPHA) daily tests are used for branching and are critical to the system. These tests could not be ommitted in a well-implemented center.

Exhibit III-17
Page 5 of 7

In a third system, PLAN, no daily testing exists for either diagnosis or student control. All testing is diagnostic, but occurs after major (1-2 week) time intervals. In still a fourth system, GINN, tests may occur daily or weekly but will not modify the reading assignment and may be given at such times when a teacher is planning new instructional skill groups.

In each system, testing is observable in the classroom. The consequences of the testing require not only that the observer interview the teacher, but also, that the observer be aware of the program developers purpose and intent for use of each of these instruments.

III.

Based on these differences, an interview form was attempted. An additional problem soon became evident. With an interview form, project differences could be determined. But, were these differences critical for the implementation of a specific project?

The following list of questions are a selection of those that quickly demonstrated differences between potential programs. Ten of these questions are listed here for purposes of demonstration. In these questions the word "module" is synonomous with Block (ALPHA), TLU (PLAN), MODULE (LU), Reading Assignment (GINN), teacher assignment for an objective (M.C.).

QUESTIONS WHICH DISCRIMINATED BETWEEN PROGRAMS OF VARIOUS CATEGORIES:

		RATING OF "TOTALLY IMPLEMENTED" CENTER				
	INTERVIEW QUESTION	ALPHA	LU	PLAN	GINN	M.C.
1.	Does student take a test which will allow for a Branching decision on a daily basis?	Yes	No	No	No	Yes
2.	Within a "module", can a student indicate objective mastery?	Yes	No	No	Yes	Yes
3.	Within a "module" can a student work on only those skills necessary to complete that module?	Yes	No	No	No	Yes
4.	Does a student usually work in more than one instructional modeule in a given area at a given time?	Yes	Yes	No	No	Yes
5.	Does system allow teachers to determine if a student is functioning in the system and accomplishing assignments successfully?	Yes	Yes	No	No	No

QUESTIONS WHICH DISCRIMINATED BETWEEN PROGRAMS OF VARIOUS CATEGORIES: (Cont'd)

INTERVIEW QUESTION	RATING OF "TOTALLY IMPLEMENTED" CENTER				
	ALPHA	LU	PLAN	GINN	M.C.
6. Does system allow daily indication of student work effort?	Yes	No	No	No	No
7. Does system provide systematic review of student performance to administration?	No	Yes	Yes	No	Yes
8. Does system provide systematic review of student performance to parent?	No	Yes	Yes	No	No
9. Does system stress teacher presentation of skill?	Yes	No	No	Yes	No
10. Does system stress use of incentives in classroom?	Yes	Yes	No	No	No

If we use these 10 items with equal weighting for each (yes=1 no=0) to evaluate the degree of implementation in project sites, the following information is obtained:

ACTUAL SAMPLE RATING OF CENTER ON TEN INTERVIEW ITEMS:

ITEM	ALPHA	LU	PLAN	GINN	M.C.
1	1	0	0	0	1
2	1	0	0	1	1
3	1	0	0	0	1
4	1	1	0	0	1
5	1	1	0	0	0
6	1	0	0	0	0
7	0	1	1	0	0
8	0	1	1	0	0
9	0	0	0	0	0
10	0	0	0	0	0
Percent of Implementation	60%	40%	20%	10%	40%

It becomes obvious with this example, that the implementation in an actual project site can not be determined by a standard implementation interview form.

Exhibit III-17
Page 7 of 7

IV.

The degree of implementation must be determined by comparing the "theoretical" for each program to the "actual" for each program site. If this is done with this particular series of project observations, a totally different profile for implementation is obtained.

TOTALLY IMPLEMENTED PROGRAMS:
V = Variable is relevant
- = Variable is not relevant

	ALPHA	LU	PLAN	GINN	M.C.
1	V	-	-	-	V
2	V	-	-	V	V
3	V	-	-	-	V
4	V	V	-	-	V
5	V	V	-	-	-
6	V	-	-	-	-
7	-	V	V	-	V
8	-	V	V	-	-
9	V	-	-	V	-
10	V	V	-	-	-

ACTUAL SITE OBSERVATIONS ON PROJECT:
(Relevant variables only)
1 = actually occurs 0 = doesn't occur

	ALPHA	LU	PLAN	GINN	M.C.
1	1	-	-	-	1
2	1	-	-	1	1
3	1	-	-	-	1
4	1	1	-	-	1
5	1	1	-	-	-
6	1	-	-	-	-
7	-	1	1	-	0
8	-	1	1	-	-
9	0	-	-	0	-
10	0	0	-	-	-
	6/8	4/5	2/2	1/2	4/5
	75%	80%	100%	50%	80%

In completing this analysis, we selected those questions which did discriminate between sites. In a total interview form, these questions would not make the dramatic difference that they did with this example. However, the difference would be great enough to invalidate any attempt to use degree of implementation as a variable which might influence program outcome.

To ensure that critical variables to show up in the final analysis, program developers will assign weights to all activities and processes in observation that they specify as a part of the implementation procedure.

A. Suggested Procedure

The same general categories of implementation activities (Exhibit III-15) which would be common to virtually all learning systems would be used to determine the degree of implementation of both individualized and standardized programs.

The process which would be used upon final selection of the candidate sites would resemble that used in the SRI Follow Through study, with the exception that responsibility for including and/or expanding specific activities and criteria for assessing these activities will be shared with the model developers or publishers. Based upon a review of existing implementation procedures and observation instruments used for monitoring implementation, specific activities would be listed for each category and tentatively weighted as "critical", "important", or "not important". Those specific activities, procedures, and assessment criteria will constitute the data requirements for instruments to be used to verify the extent to which a potential candidate is well-implemented, prior to final selection and inclusion in the study. These data requirements for instruments will be included as an addendum to the questionnaire to be sent to the appropriate LEA officials, as described in Appendix 5. Additional telephone and on-site follow-up will be required to complete the verification process regarding the critical activities and events relevant to the operation of the program during 1976.

In Appendix 11, we have identified anumber of critical events and activities for a sample potential candidate site. Using the general categories in Exhibit III-15, the candidate site staff and consultants will have listed and weighted the specific items and criteria which will be used for evaluating implementation in their site. In this example, the data requirements and assessment criteria reflect the idiosyncracies of implementation procedures inherent in this particular program.

To develop the specific requirements for implementation it will become necessary to not only involve the publisher or model developer but also the most knowledgeable individual(s) at the sites to determine the specific ways in which the programs "should be" implemented. The specific items which must be agreed upon will include: (a) an explicit description of classroom procedures (e.g., diagnosis, prescription) and alternatives which are acceptable for well-implemented programs; (b) a relative weighting assigned to each of the items or observations; (c) specific criteria indicating the accomplishment -- or lack thereof -- with respect to each of the activities or procedures (e.g., percentage of materials which are observed to be available and accessible); and (d) the identification of source documentation wherever possible.

The manner by which mutual agreement will be sought will differ to some extent from site to site and from program to program, depending upon the complexity of the implementation procedure and existing documentation. However, for the most part, it is anticipated that the project staff will specify the categories (as shown in Exhibit III-15) and contribute to the basic design characteristics of the instrument. Finalization of the instrument will require discussion and joint agreement between the project staff and the program developer at the potential site. Since it is the desire of both the project staff and the program developer to include well-implemented programs, disagreement on the instrument should be minimal.

Once the data requirements have been specified with respect to the procedures and assessment criteria for each of the programs, a determination will be made regarding the specific type of data collection instrument to be used. We anticipate that the types of instruments will include: interviews with building and district staff, audit checklists, and observations. The purpose of including items in interview forms, as well as audit checklists, will be to collect initial data and to verify this data through review of evidence. The major purpose of observation will be to verify the existence of specific activities. In addition to relatively structured interviews, opportunities for unstructured interviewing will also be available as a follow-up to observation. For example, observers will be trained to identify critical activities and events. Upon the occurrence of such activities (e.g., the teacher prescribed a specific learning activity for an individual student), the observer (at the completion of the class period) would ask the teacher why she/he assigned that learning activity to that child. The observer will not only record the teacher's response, but also identify the specific instructions in the teacher guide).

While the general procedures for integration during the overall data collection and analysis activities are described in Task 5, the implications for data analysis, utilizing this approach, are important. First, the above procedures will ensure reliable and valid data regarding the degree to which a potentially well-implemented program is actually being implemented. Through scaling procedures and the use of mutually agreed upon weights, one will be able to report upon the degree to which discrepancies occur in important implementation categories. Within a particular program it is anticipated that analysis will provide useful insights regarding the degree to which implementation occurs in specific classrooms. Second, to a much lesser extent, one will be able to identify the degree to which discrepancies occur between programs in general category areas and assess the impact of the variables as independent program variables upon dependent or outcome variables. However, it is expected that some judgmental decisions will be needed in assigning values to the expected differences in critical variables. Third, this approach would allow one to capture more relevant information on the qualtative nature of implementation activities as they vary within the program and within the classrooms. Discussions with publishers and model developers always tend to "emphasize" the qualitative nature of implementation activities as being critical. Moreover, this approach would also provide the opportunity for identifying not only the existence of barriers which impede effective implementation but also the justifications provided by respondents. While the collection of anecdotal data in this area may be difficult to quantify, the collection of such information has assisted enormously in past projects involving project team members (e.g., the USOE Project in the Use of Incentives) and explained other findings which emerged from hard data analysis.

B. Optional Procedure

As requested in the RFP and subsequently by the NIE Project Officer, we have devloped an optional procedure and instrument for assessing implementation across programs *only* with respect to those dimensions of individualization on which the programs are designed to differ. The feasibility of using this instrument will be tested early during the implementation phase. If feasible, it could be integrated into the "suggested" procedure previously described and would have the advantage of allowing for cross-program comparisons.

The primary purpose of this set of procedures and instruments would be to assess programs during their *operational* phases. Hence, all categories described in Exhibit III-15 under Program Operations would be collapsed and grouped into the four dimensions of individualization: (1) performance objectives; (2) continual diagnosis and prescription; (3) student pacing; and (4) alternative paths of instruction. Even though these dimensions are considered as discrete components in Task 2, it should be noted that operational activities in these four areas overlap to *some* extent. Moreover, since predetermined sets of implementation variables will have been defined (as in the SRI Follow Through Study), the need for greater interaction with LEA staff, and possibly developers/publishers, will be even more critical in obtaining accurate information and assigning weights to specific variables (Discussion with SRI staff indicate the time-consuming nature of these interactions.) The suggested procedures and steps to be followed by the implementation contractor are similar in some respects to the suggested procedure described above.

First, after initial selection, an on-site visit will be required to verify specific dimensions of the program at the site and to determine the relative importance of variables under each of the dimensions which are appropriate for that program. It is probable that the implementation contractor will have to meet with a number of LEA staff involved in developing the program and/or training teachers for its implementation.

Second, during the on-site visit, an instrument such as that suggested in Appendix 12, will be reviewed with the appropriate LEA staff, who will: (a) identify those variables which are not applicable (N/A) to the program; and where applicable (b) indicate the acceptable range of response (or observation); and (c) assign a weight to implementation variables. In order to reduce respondent biases, it is important that the initial responses be completed for all items prior to having knowledge that weights will be assigned. Also, in describing the purpose of the study, it should be made clear that the major focus is upon the outcomes of well-implemented programs, thereby implying the advantage of narrowing the acceptable range as closely as possible to the prescribed implementation plan.

Third, where possible, the implementation contractor's staff, while on site, will administer the instrument in a limited number of sites, in several classrooms, to determine its feasibility. It is anticipated that two or three staff members involved in the refinement of the instrument, or one similar to that in Appendix 12, will conduct these observations, insuring some degree of interrater reliability.

If the above procedure and instrument are feasible and provide accurate and reliable data, then, upon review by the implementation contractor and NIE, a decision will be made as to whether or not to use this instrument in lieu of portions of the instrument described under the suggested procedure above. The major advantage of this option is, of course, that cross-program comparisons regarding degrees of implementation could be made, possibly relating them to outcome measures. It should be noted, however, that implementation will be assessed only along those dimensions of individualization by which the programs are designed to differ. In addition, aside from information regarding planning, teacher training, and project start-up, the differences regarding implementation between individualized and standardized programs will be significant since the dimensions to be assessed are those reflected in individualized programs.

We have included the above optional procedure and instrument in direct response to a request from the NIE Project Officer. We are still, however, very concerned regarding its feasibility and utility. This feeling has been buttressed more recently in discussions with additional Directors of Follow Through models which were included in the SRI study, and other knowledgeable consultants, including Dr. Richard Rossmiller at the Wisconsin Research and Development Center for Cognitive Learning.

TASK 4 -- DEFINE OUTCOME BEHAVIORS

Given the multifaceted character of educational goals, three general categories of criterion variables will be incorporated within the design of this study. The first, academic achievement in reading and mathematics, requires no special justification since it is a generally agreed upon priority in compensatory education. In addition, while achievement in both areas is strongly related, reading scores at the 2nd and 3rd grade level tend to be indicative of achievement in all the content areas at the junior and senior high level. This point has been supported by the research of Bond and Dykstra (1967) and Neuman (1972). The two other areas to be considered in the design are: (a) student and teacher attitudes toward the school environment; and (b) student self-concept. Averch, et al., (1971) have criticized the measurement of limited school outcome variables within educational research stating that:

> "educational outcomes are almost exclusively measured by cognitive achievement. But the educational system has many functions and many outputs. Cognitive achievement, in paritcular that part measured by standardized tests, is only one aspect of student learning (p. 153)."

Both of these affective areas are seen by many as having important educational value because, as some educators believe, development of student affect is an important end in itself (e.g., Holt, 1964) and has serious implications for student success in future academic endeavors (Bloom, 1964). Such affective characteristics of both students and teachers may help to modify learning outcomes by directly affecting the amount of time available for effective learning acquisition and by affecting student motivation towards learning. These contentions are supported by the extensive research efforts in this area (e.g., Brookover, Thomas and Patterson, 1964; and Dryer and Haupt, 1966). Thus, in order to evaluate educational programs and their effectiveness for promoting student learning, the design must focus not only on students' achievement but also on the qualities of the educational environment that are though likely to promote such achievement. This approach provides the opportunity for identifying potential factors which may help to explain differential achievement across programs.

In assessing outcome variables, only those instruments which provide an objective and valid means for assessment and scoring will be considered for use. This reduces the need to use observational procedures as a means of assessing outcomes. The major arguments against observational systems and procedures to measure outcomes is that they suffer from many methological flaws, including: (a) reliability problems due to lack of interrater agreement and instability of the behavior being observed, and (b) difficulities in accurately observing the actual behaviors of interest due to the obtrusive effects which observers have on the individual and/or the environment being observed, and sometimes due to the actual nature of the variables of interest. Herbert and Attridge, as referenced in the RBS proposal, point out that the work which has thus far been done in the identification and solution of the methodological problems of observation is inadequate to ensure accuracy of the findings.

A Rationale for the Use of Standardized Achievement Tests for Assessing Academic Skills

If one were to think of individualized and standardized programs as differing only with respect to instructional procedures with exactly the same objectives, the task of comparing the effectiveness of the two instructional approaches would be relatively easy; build tests tailor-made to the programs' common objectives, administer them, and compare the results across programs. However, the very nature of individualized programs delimits the effectiveness of using a single tailor-made test since the objectives of specific programs often differ or, at the very least, may be prioritized or sequenced differently across grade levels. Thus, no single tailor-made test can possible be equally fair (or equally comprehensive or complete) for all programs.

Rather than using many specific tailor-made tests which prohibit comparison, it will be more appropriate to use a single, more general, achievement measure such as one of the standardized norm-referenced instruments. These instruments are not designed to measure a narrow set of specific objectives that might be found to exist for any given program, but are designed, according to publishers of such tests, to assess achievement in a fairly diverse area. Contrary to the writings of some uniformed critics of such general tests, they do have some content validity as they are (in many cases) designed to measure a broad "national curriculum". This "national curriculum" is identified by carefully surveying and assessing curriculum experts (as to what the curricula should contain), the diverse set of curricula found in various educational systems, and curriculum textbooks and other materials which are receiving wide use in educational institutions. There is no prior reason to believe that such tests more closely match the reading and mathematics objectives in "standardized" programs than they do the objectives in "individualized" programs. However, it is true, of course, that such tests will have differential content validity for various specific programs. The argument being made here is that there should not be a bias for (or against) individualized vs. standardized programs in general.

Another argument used against standardized instruments for program comparisons is that since standardized tests are designed to differentiate among individuals, they are not effective in differentiating among programs. Such an argument seems to assume that the individual's score is the unit of analysis and that within-group variability attenuates the probability of obtaining significant between-group variability. However, this argument is not relevant if the unit of analysis is the classroom or some larger group.

Selection of an Instrument for Assessing Achievement in Arithmetic

The process of selecting a math achievement test involved two general steps: (a) the initial identification of potential instruments to be given further consideration; and (b) the final selection of the instrument to be used.

In step (a), the following criteria were used in the instrument identification process:

- The instrument has general acceptance and wide use throughout the nation;
- Strong evidence of content validity for currently used instructional programs in math;
- The availability of at least two parallel forms of the instrument;
- The appropriateness of the instrument for use with second and third grade children;
- The time required for test administration is no more than 45 minutes per subsection;
- The procedures which are used in scoring the instrument are highly objective and efficient.

On the basis of these criteria, the following instruments were identified for further consideration:

- Metropolitan Achievement Test
- Bobbs-Merrill Arithmetic Achievement Test
- Comprehensive Test of Basic Skills - Arithmetic
- Iowa Test of Basic Skills - Arithmetic
- SRA Achievement Test - Arithmetic
- Stanford Achievement Test - Arithmetic

These initially identified instruments were then given a more in-depth assessment, using information gleaned from tests and manuals, discussions with publishers, technical reports, mental measurement yearbooks, and journal articles. In this assessment, the following factors were given major consideration in the final selection process:

- The existence of a sufficient range of achievement over which the instrument effectively discriminates to minimize the possibility of ceiling and/or floor effects;
- Adequacy of the instruments' general layout, required reading level and item formats for use with second and third grade children;
- Adequacy of interval consistency among items within each subsection of the test (where a KR-20 coefficient of .80 is the minimum criterion);
- Evidence that various forms of the instrument are parallel (including: a minimum alternate form reliability of .85, comparable item difficulities between forms, comparable numbers of items related to specific objectives between forms);
- The technical excellence in the procedures used in the generation of national norms;
- Evidence that the instrument is culturally fair;
- Evidence that the time for conducting the logistics of scoring and related activities could be minimized (i.e., less than 5 days).

After a critical review of the above instruments with respect to the these criteria, it was concluded that the 1973 Stanford Achievement Test (SAT) - Arithmetic should be used for both pre- and post-assessment of achievement in arithmetic. The specific levels, forms, and test sections of the SAT to be administered to 2nd and 3rd grade students on pre-testing and post-testing are presented in Exhibit III-18.

It should be noted that in the final review process, the SAT was not only found to be acceptable with respect to all of the criteria listed above, but also compared favorably with the other instruments under consideration. The one exception to these findings dealt with alternate form reliability for which no information was available. However, other evidence was present which supports the contention that the alternate forms are in fact parallel.

Some of the outstanding features of the SAT which provide a strong argument for its use include:

(a) The content of the SAT is compatible with the changes in curriculum and instruction which have taken place over the last decade. This outstanding asset is partially due to the recency of the development of the latest version of the SAT (i.e., 1973).

(b) The SAT provides a rather complete coverage of the major instructional objectives found in the diverse math curricula used today. This is partially made possible through the narrowing of the grade range which each level of the SAT addresses, thus making it possible to incorporate greater breadth and depth for test objectives.

(c) The SAT provides "scaled scores" obtained through the use of Thurstone's absolute scaling procedure. This has resulted in the development of a system of interbattery standard scores which permit the translation of raw scores at each level to absolute scores with comparability across test levels and forms within each section of the test. Scaled scores have the unique advantage of providing approximately equal

Exhibit III-18

LEVELS, FORMS, AND TEST SECTIONS OF SAT - ARITHMETIC TO BE ADMINISTERED

SECOND GRADERS		THIRD GRADERS	
Level and Form	Test Sections	Level and Form	Test Sections
PRE-TESTING			
Primary I Form A	Math Concepts Math Computation Math Application	Primary II Form A	Math Concepts Math Computation Math Application
POST-TESTING			
Primary II	Math Concepts Math Computation Math Application	Primary III Form B	Math Concepts Math Computation Math Application

units on a continuous scale. Thus, these scales are of vast importance in measuring achievement gains over extended periods or for use in groups of students which show extensive variability in achievement. The combined effect of points (b) and (c) is that a greater number of the test items are at an appropriate level for assessing students at a given grade without the loss of comparability of scores.

(d) The national standardization of the SAT is outstanding for this kind of instrument. The standardization sample of 275,000 students was selected to represent the national population in terms of geographic region, size of city, socioeconomic status, ethnic background, and school ownership (i.e., public vs. non-public) as well as other factors. In addition to the highly representative national norms generated in the standardization process, there are also inner-city norms available which may be more representative of the kind of population with which this study is involved. Thus, both sets of norms may be used for making appropriate normative comparisons for each of the well-implemented instructional approaches.

(e) The SAT provides an index of instructional objectives which identifies and describes the behavior assessed by each item. Through the use of this system, it is possible to obtain more meaningful comparisons between local curriculum objectives and the behavioral objectives called for in the items. For this study, such a capability is important since it will provide a means of assuring comparable content validity for standardized and individualized approaches. This will be accomplished by identifying those objectives which are compatible with specific curriculum objectives; then by comparing the average number of objectives which match this test for both the standardized and individualized programs. If they are not significantly different, it is reasonable to assume the test is program fair. If a program bias were found to exist for some of the objectives, than a separate analysis would be carried out on those objectives which are common across the two instructional approaches. (See data discussion of Additional Procedures.)

(f) The items incorporated in the SAT were carefully constructed so as to be relatively free of bias with respect to race, religion, sex, and ethnic background, while at the same time reflecting the cultural diversity of children in the U.S.

Selection of an Instrument for Assessing Achievement in Reading

The procedure followed in selecting a reading achievement test was the same as that used in arithmetic in that the same two-step approach was taken, with comparable criteria used for initial identification and final selection of instruments. Through this process, the following instruments were identified for further consideration:

- California Achievement Test - Reading
- Comprehensive Test of Basic Skills - Reading
- Gates-MacGinitie Reading Test
- Metropolitan Achievement Test - Reading
- Stanford Achievement Test - Reading

A critical review of the above instruments with respect to the criteria specified earlier resulted in no single instrument providing clear superiority over all others. For this reason, the SAT was selected for use since (a) it was deemed to be of comparable quality to the best of the other instruments in this group with respect to the criteria used, (b) it has the same additional advantages that were mentioned for SAT-Arithmetic, and (c) by using the same instrument for arithmetic and reading, there are some additional advantages which accrue, including comparability of normative standards, and a simplication of the administration and scoring process. The specific levels, forms, and test sections of the SAT to be administered to 2nd and 3rd grade students on pre-testing and post-testing are presented in Exhibit III-19. In addition, by using the SAT for assessing growth in both areas, we anticipate greater efficiency and priority in scoring, a critical consideration in this study.

Additional Procedure for Assessing Cognitive Outcome Measures

One of the major testing concerns during deliberations among members of the project team focused upon the issue of program fairness, namely, whether or not the variance in content and sequencing among programs would be so great as to preclude the use of an existing standardized norm-referenced test as the sole measure for cognitive outcomes. While this issue cannot be resolved until the specific programs to be studied have been selected, and even though the SAT appears to have wide enough coverage to ensure content validity, we felt it appropriate to describe the alternatives which the project team considered in arriving at a recommended additional approach. It should be noted that the procedures outlined below and the recommended approach are very similar to that suggested in a study effort conducted by Dr. Stephen Klein (1972).

In our review of EPIE and other critiques of individualized programs, as well as analyses of appropriate materials included in instructional packages sent by publishers, we found some variance among programs in the number of performance objectives and proficiency levels for mastery, the degree of overlap of objectives and skills to be taught, and the sequencing of content and objectives. In addition to these differences based on project experience, it was also agreed that the relative importance of objectives varied and that the amount of time to achieve specific objectives would vary among various types of students involved. In a similar manner, Klein has categorized the major factors any program comparison technique should take into account as follows:

- success in mastering objectives;
- relative importance of objectives and their overlap;
- time spent in achieving objectives;

Exhibit III-19

LEVELS, FORMS, AND TEST SECTIONS OF SAT - READING TO BE ADMINISTERED

SECOND GRADERS Level and Form	Test Sections	THIRD GRADERS Level and Form	Test Sections
PRE-TESTING			
Primary I Form A	Vocabulary Reading Comprehension I Reading Comprehension II Word Study Skills	Primary II Form A	Vocabulary Reading Comprehension I Reading Comprehension II Word Study Skills
POST-TESTING			
Primary II Form B	Vocabulary Reading Comprehension I Reading Comprehension II Word Study Skills	Primary III Form B	Vocabulary Reading Comprehension Word Study Skills

- number and kinds of studnets involved; and
- cost and resources.

Prior to reviewing Klein's work in this area, we considered various alternatives as described below.

First, in addition to administering a standardized norm-referenced test, one could also administer one of several criterion-referenced test banks (e.g., Zweig, Fountain Valley). We reviewed the following criterion testing programs: IOX, Fountain Valley, Prescriptive Learning, National Assessment Tests, and criterion testing programs embedded in the content of various programs. A critique of criterion testing programs recently conducted by the Center for the Study of Evaluation (Kosecoff, unpublished) was also very useful. Our general consensus after this review was that administration of any criterion test along with a standardized norm-referenced test would probably result in test "overkill" and lessen the degree of LEA cooperation; that none of the above tests would be suitable in their existing form and would require some redesign (i.e., selection of items); and that the cost of administering and scoring such tests given the reporting date constraint would be prohibitive.

Second, we considered the alternative of either designing or developing a simple criterion testing program for the purposes of this project. This alternative would entail additional cost and would be extremely time consuming in addition to the above limitations or disadvantages. In addition, we considered the alternative of utilizing the Cohen/Random House HILS computer-based criterion testing program. Essentially, this procedure would entail determining the specific materials used in the various programs and then requesting a computer printout on the scores and/or performance objectives covered in the programs. This alternative would require administration of criterion tests and/or program audits of individual student mastery records and verification of teacher accuracy and reliability in scoring.

A third alternative considered was the administration of two or three nationally normed standardized tests assigned randomly to students in the program/classroom; converting test items to stated performance objectives and/or skill areas; determining which performance objectives were common across programs as well as those unique to specific programs; and then assesing programs in terms of the number of "common" objectives and unique objectives mastered. The major advantages of this approach would be a lack of need for additional testing as in the case of the administration of criterion tests, and lower administration, scoring, and logistical costs.

The process of converting test items to performance objectives has been done in the Kalamazoo Public Schools and elsewhere. Based on a review of the Kalamazoo testing system (e.g., a conversion of the Metropolitan Achievement Test -- MAT) and discussions with representatives of Harcourt, Brace, and Jovanovich (publishers of the MAT), it was determined that such an approach was feasible although the number of items per objective would vary considerably. In addition, the conversion process would be extremely time consuming, although less so than developing a new criterion testing program. The procedure for matching test items (now converted to performance objectives) to specific objectives in the program could either be done through reviews with individual teachers (e.g., as in the case of Kalamazoo) or through a review of existing materials, relying upon project staff and/or computer prescriptions from the HILS system mentioned earlier. Such a materials review would be extremely time consuming. Dependence upon teachers could result in unreliable responses (e.g., a teacher might hesitate to state she taught a specific performance objective if she felt students did not achieve that objective). Review by a panel of "experts" would probably be more desirable for that reason.

A review of Klein's publication subsequent to the considerations of alternatives by project team members indicated close similarity in identifying advantages and disadvantages. For example, while Klein indicated that standardized tests are developed without any intentional bias for one program over another and would, hopefully, not favor the content of one program over another, he did argue that a standardized norm-referenced test may inadvertently favor one program over another in terms of emphasis of objectives measured; such tests might not cover all relevant objectives or even all those which overlap and would not consider the difference in the number of objectives covered. Similarly, "curriculum-embedded" tests would not allow for the handling of differences in test difficulties, number of objectives covered, relative importance of objectives, or time spent to master objectives for cross-program comparisons. Assessing the techniques of developing criterion tests for objectives common across all programs, Klein noted that the major advantage here would be a program-free test measuring growth only on common objectives; however, it would not focus upon objectives unique to each program; it would be highly dependent upon a larger degree of overlap between program objectives, and if overlap is slight, then only a small aspect of each program would be considered. Moreover, this approach would not relate the importance of objectives across programs.

The concept of "program fair" testing (Popham, 1969) would measure programs on both common objectives and idiosyncratic objectives of each program being compared. While taking into account the most critical of the six factors described earlier, as Klein notes, this procedure has no provision for systematically handling differences in the relative importance of objectives, and the number of objectives the program is attempting to achieve.

The conceptual framework for analyzing alternative approaches as well as general findings described by Klein and those arrived at independently by members of the TURNKEY project team are very similar indeed. As Klein notes, "In addition, they (program-free and program fair testing concepts) both leave unanswered many questions which must be addressed, if we are going to perform realistic comparisons of different programs. For example, what are we to do if the objectives of the programs being compared do not overlap to a high degree, or if they share no common objective? What if we are faced with a more complex situation such as when one program has a total of 25 pupils, takes all year, and has 10 objectives; while a second program has 500 pupils for a semester and has only five objectives, and only two of these overlap the objectives of the first program?" (p. 9)

In the event that a preliminary review of the programs selected for study indicates the existence of these considerations, the application of this procedure becomes an even more important aspect of the study. The proposed approach outlined below reflects a synthesis of the procedures recommended by Klein and those independently derived by the project team. It should be emphasized that the major purpose of this set of procedures is to determine the relative effectiveness of the various programs and not to measure the absolute amount of success of a specific program.

1. Select test

   Rather than selecting and/or developing criterion test systems for administration, we propose to convert the specific test items in the standard norm-referenced test (i.e., SAT) previously selected for administration to performance objectives. The SAT will report scores on individual students as a criterion test does; moreover, it provides normative data (e.g., difficulty levels) on each item. To ensure broader coverage, it might be necessary to select an additional test with similar properties (e.g., the MAT), randomly assigning each test to students in each classroom.

2. Determine relative importance of "objectives"

   By using SAT, the process by which the test was developed will ensure some general agreement upon the relative importance of the skills being assessed. However, to ensure greater acceptance of the approach taken and arrive at a higher degree of congruency, we proposed to have a person extremely knowledgeable about the content of each of the selected programs (e.g., the LEA curriculum staff) assign weights on a scale of 0 to 5 to each objective in ascending order of importance, or assign a total of 100 points to all objectives, with limits on the maximum number of points to be assigned to any objective. As an alternative to using firm and LEA representations, a panel of disinterested, objective experts could be assembled (see Klein, p. 11).

3. Determine the degree of success achieved by the program

   As Klein correctly notes, several problems exist in measuring success if different objective or norm-referenced tests are used: (a) the need to convert scores to a common scale since the number of items and methods of assessment will vary; (b) difficulty levels vary among objectives and instruments, as do "equal intervals" among items; (c) variance in instrument, reliability, and validity. By using the SAT, however, the seriousness of these problems diminishes considerably, especially in light of the fact that normative data on item difficulty can be determined.

The proposed steps for determining program "success" under this option would be to:

- Determine the number of students mastering each objective on the SAT who did not exhibit "mastery" (i.e., "get it right") on the pre-test.
- Calculate the percentage of students who had not previously mastered any given item mastering the item on the post-test.
- Multiply this percentage for each objective by the assigned (weighted) relative importance.
- Multiply this product by a "difficulty factor" for that objective obtained by subtracting the proportion of appropriate students (i.e., second grade, ninth month, etc.) in the national norm sample who responded correctly to that item from the number 1.0.
- Sum up these products over all objectives covered by the test.

Exhibit III-20 illustrates this procedure for a simplified application.

To summarize, the advantages of this program-fair approach are:

- A reliance on standardized norm-refernced test which would capitalize on the curriculum studies used in the development of such tests, the national norm both for large subtests and for individual items, the judgments made by panels of national experts on the range and scope of the items/objectives to include, as well as on all other test aspects such as reliability, validity, scoring ease, etc.;
- A reflection of the local importance assigned to the teaching of particular items in a way that for all other aspects being equal rewards the program which pursues more objectives than one which pursues less.
- A reflection of the difficulty of the objectives pursued by any given program.

The above approach differs from that proposed by Klein in several respects for reasons unique to this study. First, given his formula --

Exhibit III-20
Page 1 of 3

SIMPLE EXAMPLE APPLICATION OF THE ADDITIONAL PROGRAM COMPARISON SYSTEM

PROGRAMS: I and II
TEST: One which has 4 items; each item taken to be an objective
STUDENTS: Ten students sampled for each program

RESULTS:

PROGRAM I

Objective (Item) Number

Student	1	2	3	4
A	X	*	*	
B		*	X	*
C	*	X		
D	X	*	X	*
E	*	X		*
F	*	*	*	
G	X		X	
H	*	*		X
I	*	*	*	
J	*	X		*

(Key: X indicates "mastery" shown on pre-test; * indicates "mastery of item on post-test not previously mastered; blank indicates no mastery shown)

PROGRAM II

Objective (Item) Number

Student	1	2	3	4
A		*		*
B	*	*	*	*
C			X	*
D	*	*	*	*
E			*	*
F	*	*	*	*
G		*	*	*
H			*	*
I		*	*	
J	X	*		

(Key: X indicates "mastery" shown on pre-test; * indicates "mastery" of item in post-test not previously mastered; blank indicates no mastery shown)

Exhibit III-20
Page 2 of 3

SUMMARY OF RESULTS:

	Program I Objective Number 1	 2	 3	 4	Program II Objective Number 1	 2	 3	 4
Number students not showing previous "mastery"	7	7	7	9	9	10	9	10
Number now showing "mastery"	6	6	3	4	3	7	7	8
Ratio of latter to former	.86	.86	.43	.44	.33	.70	.78	.80
Average Percent "Mastery" in all Objectives			.65				.65	

Based on summary, above programs are equally "effective".

However, considering our other factors:

Relative Importance of Items in Each Program (in both cases the sum of the weightings over all objectives equals 1.0)

	Program I Objective Number 1	 2	 3	 4	Program II Objective Number 1	 2	 3	 4
Relative Importance	.25	.25	.25	.25	.10	.20	.20	.50

Relative Difficulty of Each Item Relative to National Norm

Objective Number	Proportion of National Sample Showing "Mastery"	"Difficulty Factor"
1	.10	.90
2	.50	.50
3	.50	.50
4	.90	.10

Exhibit III-20
Page 3 of 3

Calculation of Program-Fair Effectiveness Results

	Program I Objective Number				Program II Objective Number			
	1	2	3	4	1	2	3	4
Percentage of students who had not previously mastered the item mastering the item on post-test	.86	.86	.43	.44	.33	.70	.78	.80
Relative Importance	.25	.25	.25	.25	.10	.20	.20	.50
Difficulty Factor	.90	.50	.50	.10	.90	.50	.50	.10
Product of above 3 values	.19	.11	.05	.01	.03	.07	.08	.04
Sum over all objectives on program fair effectiveness measures			.36				.22	

Thus, on a program-fair basis, Program I appears to be the more effective program.

$$\frac{(S)}{(T)}\,\frac{(N)}{(C)} = \frac{(S)}{(T)(C/N)} = E;$$

Where S indicates success on the objectives in terms of total weighted scores, N indicates the number of pupils in the program, T indicates amount of pupil time in the program, C indicates total program cost, and E indicates program effectiveness -- we propose to delete "C/N" (program cost per pupil) for comparison purposes here. Rather, the cost factor will be included in the cost-effectiveness analysis of "absolute" rather than "relative" program outcomes and treated in more detail than it is treated by Klein (who acknowledges the somewhat superficial treatment of this factor in his formula) and with a different priority focus (e.g., comparing "mainstream" vs. "pull-out"). Second, T (time in program) has also been deleted since: (a) differences in student time in the program will be relatively similar due to the sample design; (b) where time differences occur, they will be reflected in the separate cost-effectiveness analysis since most resources consumed by a student are time-related. These factor deletions make clear the implication that program effectiveness is synonymous with success on the weighted objectives.

It should be noted that the procedure outlined here could be carried out in a number of different configurations:

- Comparisons of the number and weightings assigned to objectives by program;
- Comparisons of programs without weightings across objectives (approximating the standard scoring routine except for consideration of item difficulty).

Selection of an Instrument for Assessing Students' Self-Concept

As in the earlier mentioned selection procedures, a two-step process was used for selecting a final instrument. The criteria that were used in in identifying potential instruments for inclusion were:

- Age appropriateness of the instrument for use with 2nd and 3rd grade students (including: nature of instructions and test presentation, level of item vocabulary, and time requirements for test administration - maximum limit of 30 minutes was used);
- The procedures which are used in scoring the instrument are highly objective;
- The existence of one or more subscales related to "school self-concept."

The instruments which were identified for further consideration on the basis of the criteria were:

- Self-Concept of Ability Scale (SCAS)
- The Piers-Harris Children's Self-Concept Scale (CSCS) (revised version for grades 1-3 for USOE Longitudinal Survey, 1973)

The major reason that a larger set of instruments were not identified was that most instruments which assess self-concept are not age appropriate for 2nd and 3rd grade students.

A more indepth comparison of these two instruments was then made, giving major consideration to the following factors:

- Construct validity of the instrument giving special consideration to such things as correlational and factor analytic findings that are compatible with theories of self-concept (Wylie, 1961);
- Evidence of correlation with academic achievement;
- Evidence of internal consistency among items on each scale;
- Evidence of (short-term) test score stability;
- Evidence of high correlation with other instruments which assess self-concept.

The comparison conducted between the SCAS and the CSCS resulted in the CSCS being found to be far superior. The CSCS was also found to be acceptable with respect to all of the factors given major consideration, and thus, will be incorporated in the study for assessing self-concept. (It should be noted that the Purdue Education Research Center recommended the use of this instrument in their proposal for this study based upon the validation results from its field testing in the Indianapolis Public Schools.) However, in order for this instrument to be appropriate for use with 2nd grade students, the item must be read by an examiner. Thus, an oral presentation will be used for both 2nd and 3rd grade students with the scores obtained on all factors (on the revised form): feeling self, school self, and behaving self being considered separately.

Because the Piers-Harris CSCS instrument, particularly the revised form for grades 1-3, might not be widely known, a brief description of the development process and assessment of correlation with other variables is appropriate, as excerpted from McDaniel, et. al. (1973).

> Development of Instrument for Grades 1-3
>
> For the purposes of measuring self-concept of 1st grade children, a shortened form of the Piers-Harris Children's Self-Concept Scale was developed. To select items for the short form, data from two earlier studies conducted at Purdue were used. In the first study, the full Piers-Harris scale was administered to 413 3rd graders attending predominately black ghetto schools in Indianapolis. In the second study, the Piers-Harris was administered to 119 2nd grade white suburban students. Item analyses were computed for each subscale reported by Stanwyck (1972) and for those subscales reported by Piers-Harris (1969). Items initially chosen for the shortened scale were those which correlated at least .30 with the total and appropriate subscale totals for the two studies. These items were inspected for appropriateness of vocabulary for 1st grade children. An attempt also was made to balance the number of items representing each of the factors. The final form contained 40 items which could be grouped into three of the factors reported by Stanwyck: feeling self, 16 items; school self, 14 items; and behaving self, 10 items.
>
> Answer sheets with Yes-No response positions were prepared for primary grade children. Small pictures (e.g., a dog, fish, boot) were used as place markers for the items.

> This test was administered to 168 1st grade pupils in Indianapolis. Out of this sample, 25 answer sheets were discarded because of a larger number of omissions or obvious irregularities in responding. The following analysis, therefore, is based on the responses of 143 pupils. (The item analysis is summarized in Exhibit III-21.)
>
> Item-subscale correlations are considered more important than item-total score correlations in view of the effort to obtain independent scales.
>
> Four of the 16 items on subscale 1 fall below the criterion of .30 correlation with the subscale score. One item in each of the other subscales failed to meet this criterion. The wording of each of these items, together with thirteen items identified by examiners as containing troublesome vocabulary, were revised before inclusion in the final form . . . The results reported in this section are for the wording used in the Indianapolis testing. The reliability of the subscales appears to be too low to warrant analysis of individual scores. (p. 48) However, they are of sufficient magnitude for use in group comparisons.
>
> Correlations with Other Variables (Grade 1)
>
> The total self-concept score for 1st grade pupils was significantly correlated with achievement scores and some measures of independence and affiliative behavior. It was not correlated with selected measures obtained from peer ratings or parent questionnaires. (Exhibit III-21 presents the correlations between self-concept total score and selected variables.)
>
> The total self-concept scale score was correlated ($r = .23$, p .01) with reading achievement and ($r = .42$, p .01) with math achievement as measured by the Stanford Achievement Test. A significant correlation was found between the self-concept score and the independence subscale ($r = .45$, p .05) and the social subscale ($r = .59$, $p.$ .01) of the Pupil Information Booklet (McDaniel, et al, 1973). No significant correlations were found between self-concept and the subscales of the Peer Ratings. Self-concept was not correlated with the Maturity Demand Subscale of the Parent Questionnaire (McDaniel, et al, 1973) (p. 48)

It is felt by the TURNKEY study team that significant correlation between the total self-concept scale and reading/math achievement is in no way a drawback to the use of the instrument as intended in the study design. The existence of these significant correlation provides evidence of construct

Exhibit III-21

Item-Subscale and Item-Total Correlations (Grade 1)*

Piers-Harris Children's Self-Concept Scale Adapted for Primary Grades

Item	Item - Subscale Correlations			Total Score
	I Feeling	II School	III Behaving	
1	.30			.33
2	.32			.29
3	.34			.29
4	.34			.18
5			.45	.23
6		.32		.24
7		.41		.22
8		.44		.21
9	.29			.22
10		.53		.37
11			.54	.40
12			.43	.29
13		.42		.17
14		.55		.32
15		.27		.37
16			.14	.27
17			.55	.36
18	.18			.07
19	.36			.33
20		.47		.36
21		.63		.41
22			.51	.46
23		.47		.26
24		.54		.35
25			.45	.32
26		.32		.29
27		.37		.26
28	.45			.32
29	.39			.32
30	.41			.22
31	.27			.16
32	.29			.17
33	.30			.29
34			.52	.39
35	.43			.32
36		.42		.29
37			.37	.28
38	.33			.28
39			.43	.31
40	.47			.52
N = 143				
Means	7.64	8.81	6.40	22.85
SD	2.72	2.67	2.32	5.64
KR-20	.48	.67	.53	.73

*Reproduced from McDaniel, et al., 1973, p. 49.

validity for the self-concept instrument, at least regarding student performance in school. However, since the amount of variation in reading achievement scores explained by variations in the total self-concept scale ($r^2 = .05$; 5% explained variation) and the amount of variation in math achievement scores explained by variations in the same scale ($r^2 = .18$; 18% explained variation) are both relatively low, it is not reasonable to conclude that achievement tests are already measuring whatever the total self-concept scale is measuring. Thus, the correlations shown in Exhibit III-22 are considered appropriate.

Selection of an Instrument for Assessing Student's Attitude Toward School

Again using comparable criteria for initial selection as was used for the student self-concept instrument, the following instruments were identified:

- Attitude Toward School (ATS)
- My Class Inventory (MCI)
- School Attitude Test: Oral Form (SAT:OF)
- School Sentiment Index (SSI)

These instruments were compared with respect to their desirability for use giving major consideration to factors comparable to those used previously for final affective instrument selection. This comparison resulted in the selection of the ATS for use in this study because it was found to be acceptable with respect to all major factors with the exception of "high correlation with other instruments which assess students' attitude toward school" for which no evidence was available. In addition, the ATS was found to be superior to the other instruments considered. Scores obtained on all subscales: attitude toward school work, attitude toward teachers, and attitude toward school in general, will be considered as separate variables.

Exhibit III-22

Correlations between Self-Concept and Selected Variables (Grade 1)***

Scale	N	Correlation
Reading Achievement	138	.23**
Mathematics Achievement	137	.42**
Peer Rating Scale		
Individual Action	138	.04
Affective Relationships	138	.13
Pupil Information Booklet		
Independence	16	.45*
Social	16	.59**
Parent Questionnaire		
Maturity Demand	63	-.06

* p .05 (one tail test)
** p .01 (one tail test)
*** Reproduced from McDaniel, et al, 1973, p. 50

The item analysis for the Indianapolis data revealed no item-subscale correlations below .30, with the vast majority being above .40. The alpha estimates of reliability for this testing were .75 for attitude toward schoolwork, .79 for attitude toward teachers, and .81 for attitude toward school in general. The reliability for the total scale was .91. (The percent of students choosing each response for Form V can be found in Exhibit III-23. The item-subscale and item-total correlations are presented in Exhibit III-24.)

These data suggest that Form V of the attitude toward school measure is sufficiently well developed for use in subsequent studies. There is, however, a substantial intercorrelation among the subscales (schoolwork and teachers, .61; schoolwork and school in general, .76; and teacher and school in general, .69).

The high correlations suggest that the present instrument is essentially unidimensional in nature and may be measuring a single generalized orientation to the school environment. Whether this is an artifact of the instrument or reflects the actual nature of school attitudes cannot be determined at this time. Despite the apparent high overlap between the subscales of this instrument, there is some evidence from the correlations which follow (Exhibit III-25) that the subscales represent a degree of orthogonality.

Correlations with Other Variables (Grade 4)

It would seem logical that a child's attitude toward school would be related to several other variables. First, his general ability and achievement scores should be important, assuming that school attitude is not likely to be high if school is too difficult for the student, or if he is not doing well in the work required. Second, if a child's self-concept is markedly low or high, one might expect some generalization and projection of these perceptions onto the outer environment. Finally, it would seem that the parental aspirations for the child and the parental acceptance of the child would have a bearing on the way the child relates to both the schoolwork and the school teacher.

Low but positive correlations were found between attitude toward school and I.Q. measured here by Raven's Coloured Progressive Matrices . . .(Exhibit III-25). . . Orthogonality of the subscale is suggested by noting that the correlation between the Raven's and attitude toward schoolwork is higher than the correlations between the Raven's and the other two subscales of the attitude tests.

The highest correlations were found between attitude toward school and measures of achievement. These ranged from .28 to .43 and all were significant beyond the .001 level. Here again, correlations between the subscale attitude toward schoolwork and achievement were higher than those for the other subscales. These correlations are among the highest found in the literature. . .

Exhibit III-23

Percent of Students Choosing Each Response (Grade 4)*

Attitude Toward School

Number	A	B	C	D	E
1	31	6	15	16	32
2	17	7	8	19	49
3	14	6	19	11	49
4	28	5	7	12	48
5	15	3	12	13	54
6	30	7	22	7	32
7	24	8	22	17	28
8	22	9	14	19	36
9	14	7	14	13	51
0	12	8	16	18	46
1	13	6	33	8	38
2	31	6	8	9	42
3	27	13	18	14	28
4	36	11	10	13	29
15	24	9	11	14	41
16	20	8	10	14	47
17	28	9	15	14	33
18	53	8	5	7	25
19	13	5	8	20	54
20	17	3	8	15	56
21	25	5	22	9	38
22	44	8	8	9	32
23	11	6	16	26	41
24	18	6	13	13	48
25	20	6	12	14	47
26	52	17	9	6	15
27	23	7	23	12	34
28	9	5	6	19	60
29	26	6	14	13	39
30	14	9	22	16	38
31	22	8	21	16	33
32	27	9	17	14	30
33	24	17	14	12	33
34	26	8	14	11	40
35	10	3	13	21	52
36	27	12	14	7	39
37	32	8	18	13	28
38	40	9	12	8	29
39	7	4	9	12	66
40	29	7	28	4	29
41	21	11	9	6	53
42	21	7	12	13	47
43	23	8	13	13	42
44	33	10	14	7	33
45	12	2	12	13	59

uced from McDaniel, et al, 1973, p. 38

Exhibit III-24

Item-Subscale and Item-Total Correlations (Grade 4)*

Attitude Toward School

Item Number	Schoolwork	Teacher	School in General	Total
1			.50	.53
2	.47			.41
3		.60		.54
4			.49	.45
5	.53			.45
6		.59		.53
7			.53	.44
8			.51	.47
9	.51			.45
10	.34			.30
11		.37		.28
12		.57		.53
13	.52		.62	.58
14				.45
15		.65		.61
16			.66	.62
17		.38		.28
18			.53	.52
19	.59			.58
20		.67		.59
21		.45		.39
22			.48	.37
23			.36	.42
24	.42			.31
25			.53	.54
26			.38	.30
27	.63			.63
28		.46		.41
29			.47	.46
30		.43		.29
31		.61		.52
32	.42			.32
33			.58	.52
34	.54			.46
35	.49			.42
36		.57		.48
37			.55	.54
38	.59			.55
39			.41	.42
40		.40		.38
41	.51			.42
42			.45	.41
43	.30			.24
44			.39	.35
45		.51		.36
N = 179				
Mean	49.2	48.5	54.8	152.5
SD	10.7	11.4	13.8	31.9
Alpha Coefficient	.75	.79	.81	.91

*Reproduced from McDaniel, et al., 1973, p. 39

Exhibit III-25

Correlations Between Attitude Toward School and Selected Variables (Grade 4)[1]

Variable	N	Schoolwork	Teachers	School in General	Total
Raven's Total Score	169	.24*	.15	.15	.20*
Reading Achievement	179	.40**	.35**	.34**	.40**
Math Achievement	179	.43**	.28**	.32**	.38**
Self-Concept	175	.39**	.24**	.29**	.34**
Parental Aspiration for Child	58	-.32	-.34**	-.32	-.37*
Parental Acceptance of Child	86	.24	.24	.24	.27

* Significant at .01 level
**Significant at .001 level

[1] reproduced from McDaniel, et al., 1973, p. 40.

Below, a description of the development process and correlation with other variables, excerpted from McDaniel, et al (1973) is presented.

> ### Development of Instrument
>
> In the development of the instrument to measure attitude toward school, five successive forms of a questionnaire were designed and tested. Each form was tried in a school setting in order to (1) test the items, (2) establish administrative procedures, (3) check statistical and analytic methods and (4) provide ideas and approaches available only through direct testing.
>
> Three preliminary editions led to Form IV, a 45-item instrument with subscales for attitude toward school in general, attitude toward schoolwork and attitude toward teacher. Tests and retests of Form IV were administered during February and March of 1973 to students attending a parochial school in New Orleans, Louisana. Twenty-six students were in the 4th grade, 29 in the 5 grade, and 23 in the 6th grade.
>
> Teachers in grades 4-6 were asked to rate their students on attitude toward school. The 5th and 6th grade teachers rated their pupils on a scale from 1-100. The 4th grade teacher ranked her pupils to indicate the most positive to least positive attitude toward school. The correlations between the total score on Form IV and teacher ratings were .69 for grade 4, .60 for grade 5, and .66 for grade 6.
>
> Four-day test-retest correlations for Form IV are presented below:

	Grades		
	4	5	6
Schoolwork	.74	.86	.76
Teachers	.56	.74	.68
School in General	.83	.93	.76
Total	.86	.91	.80

> The correlations between teacher rating and total score on Form IV are quite high when compared with other studies in which teacher ratings were compared to other criteria . . . Scores on Form IV apparently measure attitudes which can be observed by teachers. The high test-retest correlations suggest score stability over a short period of time for the total score.
>
> Form V, the final form, represents a further effort to improve subscale reliability and internal consistency. Thirty-nine of the 45 items used on Form V had shown substantial item-subscale correlations in the previous pilot studies. Six new items were developed.
>
> The three subscales of Form V are attitude toward schoolwork, 14 items; attitude toward teachers, 14 items; and attitude toward school in general, 17 items. Form V was administered to 180 students in seven 4th grade classes in Indianapolis.

It is interesting to note the apparent relationship between attitude toward school and self-concept. These correlations ranged from .24 to .39 and all were significant at the .001 level. This is consistent with references asserting a relationship between self perceptions and attitudes toward the school environment. . .

Attitude toward school and parental aspiration for child were negatively related. Correlations ranged from -.37 to -.32. Items concerning achievement were used on the Parent Questionnaire. These items refer to the degree of concern the parents have for the success of their children. The negative correlations between parental achievement motivation and attitude toward school may be related to Erhlich's observation that high enthusiasm for elementary school seems to be associated with low socioeconomic status. Perhaps parents with a high level of concern tend to expect too much of their children which could result in the child's lower attitude toward school. This is an area relatively untouched by research and one that needs further investigation.

Positive correlations were found between attitude toward school and parental acceptance of child. This could be somewhat anticipated. Although none of these correlations were significant at the .01 level, all were significant at the .05 level.

Correlations were found between attitude toward school (Form V) and several other educational and psychological variables designated for the longitudinal study. (These are presented in Exhibit III-25.)

Selection of an Instrument for Measuring Teacher's Assessment of Classroom Environment

For this area, the set of criteria used in the initial identification phase were as follows:

- The procedures which are used in scoring the instrument are highly objective;
- Appropriateness for use in the assessment of grade school classrooms;
- Acceptable internal concsistency among items within each scale to be used where a KR-20 coefficient of .50 is the minimum criterion.

The instruments which were identified for further consideration on the basis of the above criteria were:

- Class Activities Questionnaire (CAQ)

- Learning Environment Inventory (LEI)
- Minnesota Teacher Attitude Inventory (MTAI)
- Organizational Climate Description Questionnaire (OLDQ)
- Organizational Climate Index (OCI)
- Purdue Teacher Opinionaire (PTO)
- Social Climate Scale (SCS)

A more indepth comparison of these instruments was then made, giving major consideration to the following factors:

- Construct validity of the instrument giving special consideration to such things as correlational and factor analytic findings that are compatible with theory incorporating classroom environment as a construct;
- Evidence of correlation with academic achievement;
- Availability of elementary school norms.

A critical review of the above instruments resulted in the selection of the OCI for use in this study. The one particular impressive characteristic of this instrument was the strong positive evidence available regarding its construct validity and its demonstrated "power" in assessing individualized programs. For example, Kelly, et al. (1973), in their study of teachers' perceptions of school climate resulting from the implementation of IGE and other studies sponsored by /I/D/E/A/ Kettering, found the OCI instrument to be fairly consistent in reporting the following findings for inner-city schools:

> Increased implementation of the IGE model in inner-city schools should: (1) lead to an improvement in the Intellectual Climate by increasing teacher interest in intellectual activities, social action, and improvement of personal and interpersonal effectiveness; (2) lead to teacher behaviors which reflected hard work and preseverance expressed in an increased committment of the needs of the school; (3) lead to increased sharing of help and expertise; (4) result in greater respect for each other's integrity and yet foster a willingness to help one another; (5) help to develop improved organization and procedural thoroughness in the planning and

> conducting of duties aimed at accomplishing the purposes of the school; and (6) create greater openness and flexibility in the day-to-day operation of the building.

Scores obtained on all factors comprising OCI -- intellectual climate, achievement standards, practicality, supportiveness, orderliness, and impulse control -- will be considered as separate variables. Finally, school climates isolated through combined factor analyses of the OCI and the Stern Activities Index were found to differentiate between grade levels and to be related to other measures such as teacher turnover.

The above validating information does suggest that the OCI is capable of distinguishing between institutional environments in ways that might be expected, given other information on the organizations studied.

TASK 5 -- DEVELOP ANALYSIS PLAN

Throughout the previous four tasks the design aspects of the proposed study have been developed. In Task 1 we described the variables to be studied and how the data for each would be obtained. In Task 2 we described the sampling plan, and in doing so we presented a number of critical analysis issues including:

- The Unit of Analysis -- a "classroom", see Exhibits III-7 and III-8.
- The Basic Design Framework -- centered in the four dimensions of our definition of individualization, mainstream vs. pull-out, and reading vs. math, see Exhibit III-9.
- Why the Proposed Study is to be Centered on Grades 2 and 3 -- the continuous nature of this grouping being consistent with the concept of individualization, the availability of appropriate instruments at these grade levels, and the somewhat similar nature of programs at these two grade levels.

In this task we will first discuss some of the general issues related to our analysis plan, issues that are related to the aspects of the design listed above. We will then discuss the specific research questions to be addressed by the proposed study. And finally, we will describe additional analyses that will be conducted beyond the basic statistical comparisons discussed to that point -- including cost-effectiveness analysis similar to the work done by TURNKEY in the development of the Michigan Cost-Effectiveness Model for Compensatory Education Programs.

Before discussing the specific research questions to be addressed, a number of design features should be discussed further. The first of these features to be explored beyond the discussion presented under previous tasks is the grade 2 and 3 focus of the proposed study. It was mentioned in Task 2 discussion that if the grades studied were not contiguous, comparisons between individualized and standardized programs at the same grade levels might be biased against standardized programs because of differences in material availability, rather than true differences in the effectiveness of one instructional approach vs. the other. From an analytical point of view, it is also suggested that for the purposes of analysis all results should be discussed and studied in terms of grades 2 and 3 combined, both for the individualized programs as well as for the standardized ones. By placing an emphasis on the combined results at these two contiguous levels, this bias problem is minimized. Separate analyses by grade level could always be generated if a comparison of results by a category such as the nominal grade level associated with a given program for administrative purposes is considered desirable. The desire to use combined results for these grade levels in our analyses can be readily accommodated within our sampling plan as already indicated by placing a constraint on the overall plan of having to include about the same number of nominally second grade classrooms and nominally third grade classrooms in each sampling cell.

The next design feature to be addressed here is the nature of the basic design framework discussed previously in Task 1 and in more detail in Task 2. To the casual reader it may seem that the Definitional Matrix (see Exhibit III-9) relied upon in this basic framework differs in concept and approach from the steps outlined in our original Proposal for developing an operational definition of an individualized or a standardized program. In our Proposal we suggested that we assign ratings to the characteristics that would make up our definition and sum up these ratings to obtain some scaled value that would reflect degree of individualization. We stated then, and believe now, that such a scale is continuous in nature and that the obvious intent of the study requested by NIE to be designed in this effort was to contrast programs from opposite ends of that spectrum in order to assess the effectiveness of individualization. The definitional matrix relied upon here is identical in concept to this anticipated scaling approach though different in appearance. If one assigns a value of +1 to each "yes" associated with one of the key issues reflected in that matrix and 0 with each "no", the scaled total ratings for the four study groups would be as follows:

	Total Scaled Rating
• Group I -- "yes" to all four issues.............	4
• Group II -- "yes" to any three issues...........	3
• Group III -- "no" to any three issues...........	1
• Group IV -- "no" to all four issues.............	0

Thus, the four study groups may be viewed as falling along a continuum of individualization and the omission of programs with total scaled ratings of 2 is consistent with our original concept of NIE's desire to contrast programs from opposite ends of that continuum. The matrix display technique was chosen because it illustrates the factorial nature of our approach better than a rating scale would, and it is felt that this factorial (or quasi-factorial) nature in the compression of the overall 16-cell matrix to the four study groups, allows much to be done in terms of powerful analytical techniques.

A point made earlier in the discussion of this definitional matrix in Task 2 warrants re-emphasis here. The TURNKEY study team has concluded that based on the best information at hand and knowledgeable estimates of anticipated information in this regard, the shaded area of the definitional matrix would be excluded from the analysis. This decision is obviously in accord with the discussion of the above paragraph. However, in the interest of thoroughness it was pointed out in Task 2 that it may be that filling in all cells of the definitional matrix is feasible -- a fact that might become apparent to the implementation contractor as further candidate sites are identified. If this is the case, in spite of any initial desire to observe programs only from the ends of the individualization continuum rather than from the middle ground, from a factorial design point of view it may be desirable to study such programs in order to more fully investigate higher level interactions among the four key issues related in the definitional matrix. This deviation from the analysis plan suggested here is an option that depends entirely on the availability of sufficient data for such middle ground programs -- the probability of which is assumed to be low by the TURNKEY study team at this time.

Research Questions to be Addressed

For ease of reference the basic design framework described earlier in detail in Task 2 is presented again here as Exhibit III-21. The individual cells are identified by the numbers 1 to 16; the four study groups of programs are identified as numbers I to IV. Exhibit III-22 repeats the specific four sampling frames -- or repetitions of the basic framework iterated over the other sampling dimensions (mainstream vs. pull-out and reading vs. math) besides individualization/standardization.

This design framework allows a number of research questions to be addressed. The discussion below presents these questions within a hierarchial structure of 4 levels. The Level 1 issues are those most basic to the proposed study. Levels 2, 3, and 4 include issues that are important to address in the study, but the critical value of the issues addressed (and ability of the proposed design to provide powerful results regarding these issues) decreases at each step along this hierarchy.

The questions presented below are to be answered by means of a series of contrasts. The specific data to be contrasted in order to answer each specific issue could be any, if not all, of the dependent variables presented in an earlier section, excluding the "other" variables discussed at that time. Included in these possibilities is the additional procedure for measuring program-fair success discussed in detail in Task 4. In each case the results that are compared will be the combined results for that dependent variable from all second and third grade classrooms included in that specific sampling cell. Thus, the entire set of contrasts presented

Exhibit III-21

BASIC DESIGN FRAMEWORK

		SPECIFIC PERFORMANCE OBJECTIVES ARE ASSIGNED TO EACH STUDENT		SPECIFIC PERFORMANCE OBJECTIVES ARE NOT ASSIGNED TO EACH STUDENT	
		DIAGNOSIS/ PRESCRIPTION IS CONTINUOUS	DIAGNOSIS/ PRESCRIPTION IS NOT CONTINUOUS	DIAGNOSIS/ PRESCRIPTION IS CONTINUOUS	DIAGNOSIS/ PRESCRIPTION IS NOT CONTINUOUS
INDIVIDUAL LEARNER OR MATERIAL PATHS DO EXIST	INDIVIDUAL PACING EXISTS	1 I	2 II	3	4
	INDIVIDUAL PACING DOES NOT EXIST	5 II	6	7	8
INDIVIDUAL LEARNER OR MATERIAL PATHS DO NOT EXIST	INDIVIDUAL PACING EXISTS	9	10	11	12 III
	INDIVIDUAL PACING DOES NOT EXIST	13	14	15 III	16 IV

EXHIBIT III-22

SPECIFIC DESIGN FRAMEWORK

A. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Mainstream
SUBJECT AREA: Reading
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I (n=51), II (51), III (51), and IV (51)

B. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Pull-Out
SUBJECT AREA: Reading
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I (n=49), II (49), III (49), and IV (49)

C. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Mainstream
SUBJECT AREA: Math
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I (n=51), II (51), III (51), and IV (51)

D. GRADES: 2 and 3 Combined
NOMINAL COMP ED INSTRUCTIONAL SETTING: Pull-Out
SUBJECT AREA: Math
UNIT OF ANALYSIS: "Classroom"
PROGRAMS STUDIED: Groups I (n=49), II (49), III (49), and IV (49)

below may be repeated for each of the dependent variables for which adequate results are obtained and which are considered relevant to the questions being studied. However, given the multi-faceted nature of the outcome measures and instruments described earlier in Task 4, a listing of these measures in the order of their anticipated importance for the contrasts outlined here is appropriate. This priority listing is as follows:

1. Total scores, achievement tests
   - Reading
   - Math
2. "Program-fair" measures of effectiveness based on achievement scores
   - Reading
   - Math
3. Total scale scores, effective instruments
   - Piers-Harris
   - ATS
   - OCI
4. Sub-scale scores, achievement tests
   - Reading
   - Math
5. Sub-scale scores, effective instruments
   - Piers-Harris
   - ATS
   - OCI

The specific means or statistical techniques used to assess each contrast presented below would range from simple comparison of means or frequencies via t-test or contingency table methods, where appropriate, to more sophisticated approaches such as ANOVA or discriminant function analysis. Where specific cells or groups of cells from the definitional matrix are contrasted, techniques for unequal sample size treatment will be applied as needed.

Level 1 Analysis is presented as Exhibit III-23. Level 2 Analysis is presented as Exhibit III-24. Level 3 Analysis is presented as Exhibit III-25. Level 4 Analysis is presented as Exhibit III-26. The approximate number of classrooms involved in each contrast are also shown. Note that where these contrasts involve individual cells in the 16-cell Definitional Matrix other than cells 1 and 16, the number of classrooms indicated is strictly an approximation based on an assumption of a uniform distribution of classrooms within study groups II and III in this matrix. This assumption may be faulty given the preceding discussion on this issue, but as a means for illustrating the scope of the contrasts, it is felt that the assumption is acceptable in this case.

Additional Analysis

Beyond the detailed lists of contrasts described in Exhibits III-23 through III-26, a number of potentially critical analyses will be addressed under the heading of Level 4 Analysis. By far the most critical -- and for now identifable -- of these additional analyses is the opportunity to cross validate the results of TURNKEY's efforts in comp ed cost effectiveness analyses conducted elsewhere, specifically in Michigan. This possibility

EXHIBIT III-23

LEVEL 1 ANALYSIS

What is the impact of individualization on comp ed program effectiveness?

Relevant Contrasts:

(All programs studied)	1. Group I (A, B, C, and D combined, n=200) vs. Group IV (A, B, C, and D combined, n = 200) 2. (Group I (A, B, C, and D combined) plus Group II (A, B, C, and D combined, n=400) vs. (Group III (A, B, C, and D combined) plus Group IV (A, B, C, and D combined, n=400) 3. Group II (A, B, C, and D combined, n=200) vs. Group III (A, B, C, and D combined, n=200)
(Reading Programs only)	4. Group I (A and B combined, n=100) vs. Group IV (A and B combined, n=100) 5. (Group I (A and B combined) plus Group II (A and B combined), n=200) vs. (Group III (A and B combined) plus Group IV (A and B combined), n=200). 6. Group II (A and B combined, n=100) vs. Group III (A and B combined, n=100)
(Math Programs only)	7. Group I (C and D combined, n=100) vs. Group IV (C and D combined, n=100) 8. (Group I (C and D combined) plus Group II (C and D Combined, n=200) vs. (Group III (C and D combined) plus Group IV (C and D combined, n=200) 9. Group II (C and D combined, n=100) vs. Group III (C and D combined, n=100)

EXHIBIT III-24
Page 1 of 3

LEVEL 2 ANALYSIS

I. What is the impact of each of the specific processes included in the study's definition of individualization on comp ed program effectiveness?

- Performance Objectives?

Relevant Contrasts (recalling the 16 cells in the definitional matrix and the specific cells included in each of the four study groups of programs):

(All programs studied)
1. Cell 1 (A, B, C, and D combined, n=200) vs. Cell 3 (A, B, C, and D combined, n=50)
2. Cell 16 (A, B, C, AND D combined, n=200) vs. Cell 14 (A, B, C, and D combined, n=50)

(Reading Programs only)
3. Cell 1 (A and B combined, n=100) vs. Cell 3 (A and B combined, n=25)
4. Cell 16 (A and B combined, n=100) vs. Cell 14 (A and B combined, n=25)

(Math Programs only)
5. Cell 1 (C and D combined, n=100) vs. Cell 3 (C and D combined, n=25)
6. Cell 16 (C and D combined, n=200) vs. Cell 14 (C and D combined, n=25)

- Continuous Diagnosis/Prescription?

Relevant Contrasts:

(All programs studied)
1. Cell 1 (A, B, C, and D combined, n=200) vs. Cell 2 (A, B, C, and D combined, n=50)
2. Cell 16 (A, B, C, and D combined, n=200) bs. Cell 15 (A, B, C, and D combined, n=50)

(Reading Programs only)
3. Cell 1 (A and B combined, n=100) vs. Cell 2 (A and B combined, n=25)
4. Cell 16 (A and B combined, n=100) vs. Cell 15 (A and B combined, n=25)

(Math Programs only)
5. Cell 1 (C and D combined, n=100) vs. Cell 2 (C and D combined)
6. Cell 16 (C and D combined, n=100) vs. Cell 15 (C and D combined, n=25)

EXHIBIT III-24
Page 2 of 3

- Alternate Learner Paths?

Relevant Contrasts:

(All programs studied)

1. Cell 1 (A, B, C, and D combined, n=200) vs.
   Cell 9 (A, B, C, and D combined, n=50)
2. Cell 16 (A, B, C, and D combined, n=200) vs.
   Cell 8 (A, B, C, and D combined, n=50)

(Reading Programs only)

3. Cell 1 (A and B combined, n=200) vs.
   Cell 9 (A and B combined, n=25)
4. Cell 16 (A and B combined, n=200) vs.
   Cell 8 (A and B combined, n = 25)

(Math Programs only)

5. Cell 1 (C and D combined, n=100) vs.
   Cell 9 (C and D combined, n=25)
6. Cell 16 (C and D combined, n = 100) vs.
   Cell 8 (C and D combined, n=25)

- Pacing?

Relevant Contrasts:

(All programs studied)

1. Cell (A, B, C, and D combined, n=200) vs.
   Cell 5 (A, B, C, and D combined, n=50)
2. Cell 16 (A, B, C, and D combined, n=200) vs.
   Cell 12 (A, B, C, and D combined, n=50)

(Reading Programs only)

3. Cell 1 (A and B combined, n=100) vs.
   Cell 5 (A and B combined, n=25)
4. Cell 16 (A and B combined, n=100) vs.
   Cell 12 (A and B combined, n=25)

(Math Programs only)

5. Cell 1 (C and D combined, n=100) vs.
   Cell 5 (C and D combined, n=25)
6. Cell 16 (C and D combined, n=100) vs.
   Cell 12 (C and D combined, n=25)

II. What is the impact of having all but one of these processes present (no matter which one is lacking) on comp ed program effectiveness?

Relevant Contrasts:

(All Programs studied)

1. Group I (A, B, C, and D combined, n=200) vs.
   Group II (A, B, C, and D combined, n=200)

(Reading Programs only) 2. Group I (A and B combined, n=100) vs. Group II (A and B combined, n=100)

(Math Programs only) 3. Group I (C and D combined, n=100) vs. Group II (C and D combined, n=100)

III. What is the impact of lacking all but one of these processes (no matter which three are lacking) on comp ed program effectiveness?

Relevant Contrasts:

(All Programs studied) 1. Group III (A, B, C, and D combined, n=200) vs. Group IV (A, B, C, and D combined, n=200)

(Reading Programs only) 2. Group III (A and B combined, n=100) vs. Group IV (A and B combined, n=100)

(Math Programs only) 3. Group III (C and D combined, n=100) vs. Group IV (C and D combined, n=100)

EXHIBIT III-25

LEVEL 3 ANALYSIS

What is the impact of classroom setting (mainstream vs. pull-out) on comp ed program effectiveness?

Relevant Contrasts:

(Reading and math combined)	1.	Group I (A and C combined, n=102) vs. Group I (B and D combined, n=98)
(Reading only)	2.	Group I (A, n=51) vs. Group I (B, n=49)
(Math only)	3.	Group I (C, n=51) vs. Group I (D, n=49)
(Reading and math combined)	4.	Group II (A and C combined, n=102) vs. Group II (B and D combined, n=98)
(Reading only)	5.	Group II (A, n=51) vs. Group II (B, n=49)
(Math only)	6.	Group II (C, n=51) vs. Group II (D, n=49)
(Reading and math combined)	7.	Group III (A and C combined, n=102) vs. Group III (B and D combined, n=98)
(Reading only)	8.	Group III (A, n=51) vs. Group III (B, n=49)
(Math only)	9.	Group III (C, n=51) vs. Group III (d, n=49)
(Reading and math combined)	10.	Group IV (A and C combined, n=102) vs. Group IV (B and D combined, n=98)
(Reading only)	11.	Group IV (A, n=51) vs. Group IV (B, n=49)
(Math only)	12.	Group IV (C, n=51) vs. Group IV (D, n=49)

EXHIBIT III-26

LEVEL 4 ANALYSIS

What is the impact of other program factors beyond the processes and settings treated in Level 2 or Level 3 analyses on comp ed program effectiveness?

Level 4 analyses are to be of a general exploratory nature using data gathered on either the covariables or the "other" variables discussed in the initial section. Potential relationships between observed data for these variables and program effectiveness will be investigated using the appropriate available techniques. Specific subsamples of very successful Group I or II programs and very unsuccessful such programs could be selected and contrasts then be drawn on the values of the variables being explored between these two groups. Similar samples could be drawn from the programs in Groups III and IV. In this way, the specific classroom or program management characteristics or student factors that seem to be associated with success in each of the types of programs could be identified. It is recognized that such an exploratory analysis would at best only specify a list of factors to be consolidated in subsequent research, but if this list is generated by the proposed study, the effect would be well worth even this interim accomplishment. This particular option will be explained in more detail in this section.

was discussed briefly in Task 1 where it was mentioned that, although such an approach would be basically exploratory in nature and results would demand cross validation elsewhere if it were approached in a strictly exploratory manner, by viewing the cost-effectiveness results to date, derived from application of the same TURNKEY instruments suggested for use here, this particular analysis could actually serve as a cross-validation effort of these existing results. TURNKEY's work on the Michigan Cost-Effectiveness Model will have a series of results available for consideration in such a cross-validation effort well before the implementation of the study proposed here, thus facilitating the finalization of this specific substudy once the implementation contractor begins work on this study.

The effectiveness portion of this subanalysis would likely take the form of examining the relationship between program success and the various program characteristics that evolve from the current TURNKEY work in Michigan to determine whether these same relationships obtained on the sample of programs examined in this study. Perhaps even the path models developed in the current TURNKEY work could be examined for cross-validation purposes within the NIE sample, taking this effectiveness component beyond the level of more simple contrasts.

The cost portion of this subanalysis would involve an application of TURNKEY's widely used cost analysis methods, based on the COST-ED^{TM} Model. Exhibit III-27 displays the basic calculation stream involved in this analytical method; Exhibit III-28 is an example of the specific program cost results obtained from a comp ed program similar to those to be studied here; and Exhibit III-29 is a further example of the cost analysis results

Exhibit III-27

COST ANALYSIS METHODOLOGY

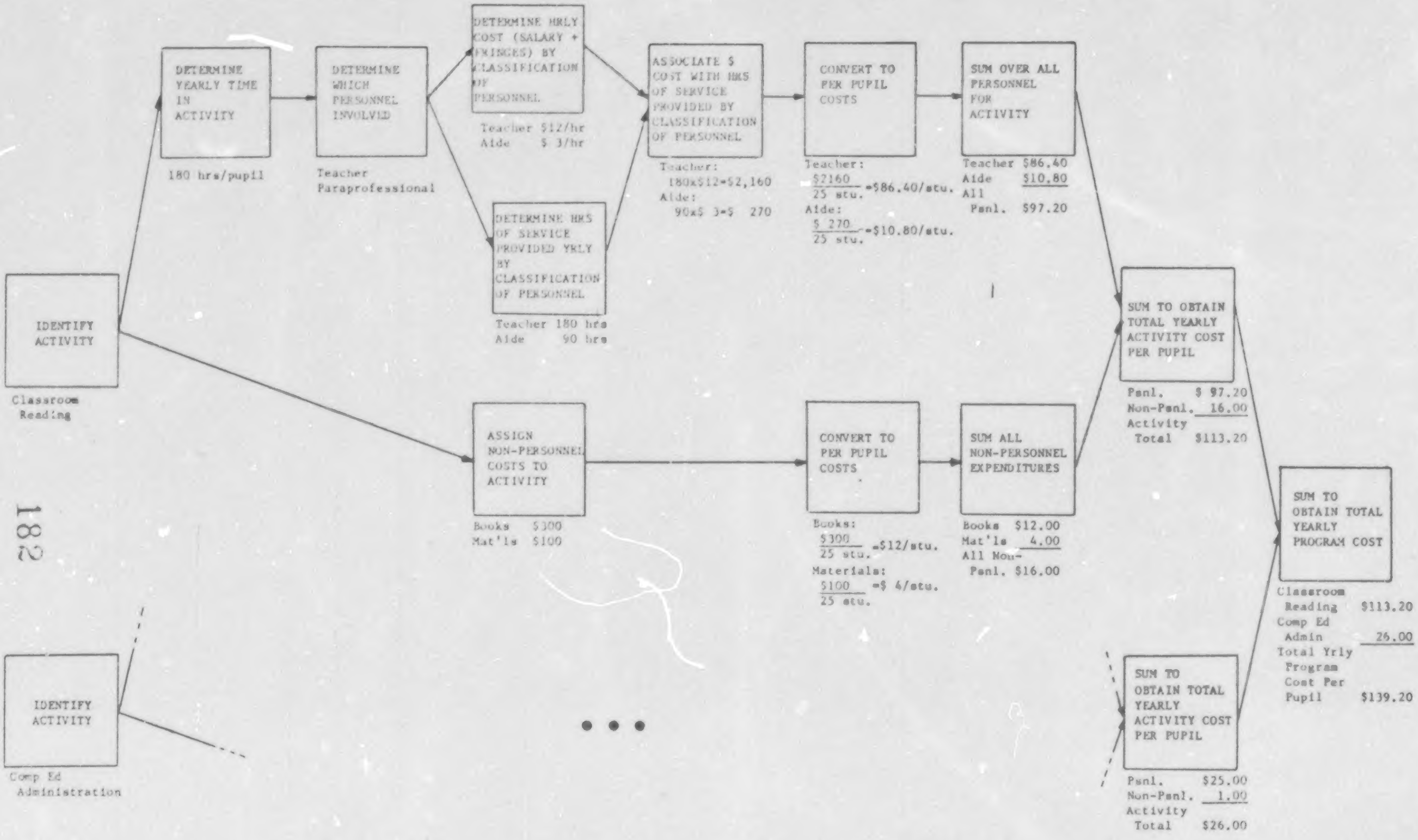

IDENTIFY ACTIVITY
Classroom Reading
DETERMINE YEARLY TIME IN ACTIVITY
180 hrs/pupil
DETERMINE WHICH PERSONNEL INVOLVED
Teacher
Paraprofessional
DETERMINE HRLY COST (SALARY + FRINGES) BY CLASSIFICATION OF PERSONNEL
Teacher $12/hr
Aide $ 3/hr
DETERMINE HRS OF SERVICE PROVIDED YRLY BY CLASSIFICATION OF PERSONNEL
Teacher 180 hrs
Aide 90 hrs
ASSOCIATE $ COST WITH HRS OF SERVICE PROVIDED BY CLASSIFICATION OF PERSONNEL
Teacher:
180x$12=$2,160
Aide:
90x$ 3=$ 270
CONVERT TO PER PUPIL COSTS
Teacher:
$2160 / 25 stu. =$86.40/stu.
Aide:
$ 270 / 25 stu. =$10.80/stu.
SUM OVER ALL PERSONNEL FOR ACTIVITY
Teacher $86.40
Aide $10.80
All Psnl. $97.20
ASSIGN NON-PERSONNEL COSTS TO ACTIVITY
Books $300
Mat'ls $100
CONVERT TO PER PUPIL COSTS
Books:
$300 / 25 stu. =$12/stu.
Materials:
$100 / 25 stu. =$ 4/stu.
SUM ALL NON-PERSONNEL EXPENDITURES
Books $12.00
Mat'ls 4.00
All Non-Psnl. $16.00
SUM TO OBTAIN TOTAL YEARLY ACTIVITY COST PER PUPIL
Psnl. $ 97.20
Non-Psnl. 16.00
Activity Total $113.20
IDENTIFY ACTIVITY
Comp Ed Administration
• • •
SUM TO OBTAIN TOTAL YEARLY ACTIVITY COST PER PUPIL
Psnl. $25.00
Non-Psnl. 1.00
Activity Total $26.00
SUM TO OBTAIN TOTAL YEARLY PROGRAM COST
Classroom Reading $113.20
Comp Ed Admin 26.00
Total Yrly Program Cost Per Pupil $139.20

EXHIBIT III-28

DOLLARS PER COMP-ED STUDENT ANNUALLY

SITE 1116 RESOURCES	FUNCTIONS					Resource Total	Percent of Total Cost
	COMP-ED READING	COMP-ED PLANNING	COMP-ED TRAINING	COMP-ED DECISION MAKING	COMP-ED ADMINISTRATION		
PERSONNEL							
District Comp-Ed Director	--	11	--	12	12	35	6.3
Principal	--	23	5	34	2	64	11.4
Comp-Ed Teacher	92	21	4	4	--	121*	21.6
Regular Teacher	63	151	--	7	--	221	39.5
Paraprofessional	82	--	--	--	--	82**	14.7
Reading Specialist	--	--	--	--	--	--	----
Other Classroom Staff	--	--	--	--	--	--	----
CONSUMABLES							
Comp-Ed Books and AV Software	27	--	--	--	--	27**	4.8
Regular Books and AV Software	6	--	--	--	--	6	1.1
EQUIPMENT							
Comp-Ed Av Equipment	--	--	--	--	--	--	----
Other Comp-Ed Instructional Equipment	--	--	--	--	--	--	----
Regular AV Equipment	--	--	--	--	--	--	----
Other Instructional Equipment	1	--	--	--	--	1	0.2
Comp-Ed Administration Equipment	--	--	--	--	--	--	----
MISCELLANEOUS							
Miscellaneous Comp-Ed Training Expenses	--	--	--	--	--	--	----
Miscellaneous Comp-Ed Administrative Expenses	--	--	--	--	2	2**	0.4
FUNCTION TOTAL	271	206	9	57	16	559	100.0
PERCENT OF TOTAL COST	48.5	36.9	1.6	10.2	2.9	100.1	

* All or part of these totals provided from Comp-Ed Budget
**All of these totals provided from Comp-Ed Budget

LEVEL: ELEMENTARY
PROGRAM: TITLE I READING

1 1 1 6
COST-ED ECONOMIC FACTOR RANKING
EDUCATION TURNKEY SYSTEMS

DATE: 09/24/74
PAGE 1

RANK OF 44	COST FACTOR	DATA GROUP	REFER.	COST FACTOR VALUE FOR CHANGE IN COST PER STUDENT YEAR: 1% SAVINGS	INITIAL VALUE	1% ADDL COST	RELATIVE POWER
1	ANNUAL SALARY	REGULAR TEACHER		7992.5905 $	8200.0000 $	8407.0156 $	100
2	STAFF RATIO	COMP-ED PLANNING	R-TCHR	7.6020 :1	7.3330 :1	7.0824 :1	79
3	STAFF USE %	COMP-ED PLANNING	R-TCHR	12.2254 %	12.7000 %	13.1705 %	68
4	ANNUAL SALARY	COMP-ED TEACHER		9418.6289 $	9875.0000 $	10331.3672 $	55
5	STAFF RATIO	COMP-ED READING	C-TCHR	5.5455 :1	5.4910 :1	5.1771 :1	46
6	STAFF USE %	COMP-ED READING	C-TCHR	20.6657 %	21.9959 %	23.3362 %	42
7	STAFF RATIO	COMP-ED READING	PARAPR	3.5365 :1	3.2940 :1	3.0827 :1	41
8	STAFF USE %	COMP-ED READING	PARAPR	20.4916 %	21.5545 %	23.5043 %	37
9	ANNUAL SALARY	PARAPROFESSIONAL		2696.5144 $	2895.0000 $	3093.4856 $	37
10	STAFF RATIO	COMP-ED PLANNING	PRINCP	85.9682 :1	78.0000 :1	71.8033 :1	33
11	STAFF RATIO	COMP-ED READING	R-TCHR	24.7242 :1	22.5450 :1	20.7191 :1	32
12	ANNUAL SALARY	PRINCIPAL		15726.4023 $	17212.0000 $	18697.5938 $	29
13	STAFF USE %	COMP-ED READING	R-TCHR	71.1252 %	77.9955 %	84.8747 %	29
14	STAFF RATIO	C.E. DEC. MAKING	PRINCP	92.9805 :1	78.0000 :1	67.1781 :1	17
15	STAFF RATIO	COMP-ED PLANNING	CE-DIR	100.3280 :1	84.0000 :1	72.2441 :1	17
16	STAFF USE %	C.E. DEC. MAKING	PRINCP	12.5842 %	15.0000 %	17.4167 %	16
17	ANNUAL SALARY	DIST.COMP-ED DIR		15363.4844 $	18350.0000 $	21336.5117 $	16
18	STAFF RATIO	COMP-ED PLANNING	C-TCHR	98.0412 :1	75.0000 :1	66.1536 :1	14
19	HR-AV HOUR COST	COMP-ED READING		0.0617 $	0.0781 $	0.0945 $	12
20	STAFF USE %	COMP-ED PLANNING	PRINCP	7.5931 %	10.0000 %	12.4169 %	10
21	STAFF USE %	COMP-ED PLANNING	C-TCHR	10.7432 %	14.8000 %	18.0357 %	9
22	FRINGE RATE	REGULAR TEACHER		4.2011 %	6.9000 %	9.5988 %	6
23	STAFF RATIO	C.E. DEC. MAKING	CE-DIR	164.1356 :1	84.0000 :1	56.4454 :1	6
23	STAFF RATIO	COMP-ED ADMIN.	CE-DIR	164.1322 :1	84.0000 :1	56.4453 :1	6
25	FRINGE RATE	COMP-ED TEACHER		6.2512 %	11.4000 %	16.5488 %	6
26	STAFF USE %	C.E. DEC. MAKING	CE-DIR	2.5585 %	5.0000 %	7.4415 %	5
26	STAFF USE %	COMP-ED PLANNING	CE-DIR	2.5585 %	5.0000 %	7.4415 %	5
28	STAFF USE %	COMP-ED ADMIN.	CE-DIR	2.5583 %	5.0000 %	7.4417 %	5
29	FRINGE RATE	PARAPROFESSIONAL		3.7623 %	11.4000 %	19.0375 %	4
30	STAFF RATIO	C.E. DEC. MAKING	R-TCHR	33.9957 :1	7.3330 :1	4.1103 :1	4
31	STAFF USE %	C.E. DEC. MAKING	R-TCHR	0.1294 %	0.6000 %	1.0706 %	3
32	STAFF RATIO	COMP-ED TRAINING	PRINCP	HIGH	78.0000 :1	35.3250 :1	2
33	STAFF USE %	COMP-ED TRAINING	PRINCP	LOW	7.0000 %	4.4193 %	2
34	STAFF RATIO	COMP-ED TRAINING	C-TCHR	HIGH	79.0000 :1	33.6703 :1	2
34	STAFF RATIO	C.E. DEC. MAKING	C-TCHR	HIGH	79.0000 :1	33.6642 :1	2
36	STAFF USE %	C.E. DEC. MAKING	C-TCHR	LOW	3.0000 %	7.0397 %	2
36	STAFF USE %	COMP-ED TRAINING	C-TCHR	LOW	3.0000 %	7.0406 %	2
39	FRINGE RATE	PRINCIPAL		LOW	4.9000 %	13.9631 %	1
39	STAFF RATIO	COMP-ED ADMIN.	PRINCP	HIGH	78.0000 :1	22.0354 :1	1
40	OTHER HOUR COST	COMP-ED ADMIN.		LOW	0.0015 $	0.0050 $	1
41	STAFF USE %	COMP-ED ADMIN.	PRINCP	LOW	1.0000 %	3.4160 %	1
42	FRINGE RATE	DIST.COMP-ED DIR		HIGH	4.9000 %	21.3940 %	1
43	USEFUL LIFE	OTHER INST. EQPT		LOW	1.0000 YRS	0.0870 YRS	0
44	RAW UNIT POMTS	OTHER INST. EQPT		LOW	1.6300 UNIT	18.7303 UNIT	0

Exhibit III-29

available from the application of these techniques -- in this case the COST-ED report known as the Economic Factor Ranking. This Ranking lists the cost factors which are part of the cost structure of any given program in the order of their impact on overall program cost, with those items ranked at the top of the page having the greatest impact and those near the bottom, the least. Such reports as shown in Exhibit III-28 and Exhibit III-29 would be available for each program/classroom studied allowing these cost results to be contrasted along with the outcome results to be analyzed as described above. Further, such cost results could even be considered together with the program-fair scheme for program comparisons described in Task 4.

A major consideration to be addressed in the cost portion of this subanalysis is the issue of standard pricing or local pricing. In question here is whether to adjust prices for various educational resources (e.g., teacher salary) to a common standard for the nation or for some other regional unit larger than the LEA. A teacher with a M.A. degree and six years of experience, for instance, will receive a vastly different salary if he worked in a large urban center in the Northeast than if he worked in a small rural district in Texas. Differences in the cost of living, the impact of local labor supply/demand conditions, and the power of local teachers organizations are just a few of many reasons for such differences. In a study such as that proposed here that is national in scope, one could argue for the use of standard pricing so that differences in costs accurately reflect differences in the amounts of resources consumed rather than simply differences in prices.

On the other hand, local price/salary structures undoubtedly influence the actual mix of resources found in any given program. The trade-offs among educational resources at two different sites using ostensibly the same program will vary depending upon the relative price of these resources; where aides are relatively less costly with respect to certified teachers, more aides will be used and where the opposite is true regarding relative prices the opposite may well be true regarding the reliance on aides. Given this interaction of the local price structure with the actual configuration of the program as operated locally and the resource mix nature of individualized instruction in general, a stronger argument could be made for the use of local or actual prices in the cost analysis.

COST-ED can be used in either a local price mode or a standard price mode so that, whichever argument is found to be most pursuasive,COST-ED analysis can readily accommodate either option. A third option -- that of computing both local and standard price-based costs is also readily accommodated by COST-ED. Other resource/cost models currently in use (e.g., the Haggert Model, the RMC Model) can be exercised more easily in the standard price mode than in the local price mode.

Beyond its flexibility regarding the issue of standard pricing, COST-ED lends itself to the cost portion of this subanalysis for a number of other reasons:

- It is flexible and adaptable to any type of instructional program (e.g., reading, math), classroom setting (e.g., mainstream, pull-out), and program organizational level (e.g., classroom, school, district-wide)

- It directly reflects the impact of the length of the program (i.e., number of hours of instruction per student) on the cost of the program.
- It provides opportunities for conducting trade-off analyses of various configurations of a specific program.
- It is useful to analysts in associating costs to program features found to be related to program success.
- It is a potentially useful incentive for LEA cooperation, since the model can provide useful data to local managers through sensitivity analyses. Such analyses can help in identifying alternative ways to cut costs if the manager is faced with budget cutbacks or in allocating additional funds if the opposite is true.

Treatment of Implementation Results

It is suggested that implementation data be used in a number of ways. First of all, those classrooms not considered well implemented (though few are expected to fall through the sample selection process described earlier) will be dropped from the contrasts that are included in Levels 1, 2, and 3. This action will undoubtedly result in unequal n contrasts in most if not all cases; appropriate techniques to reflect these unequal n situations will be applied as needed.

Second, the impact of implementation considerations on program effectiveness could be assessed by considering the differences in effectiveness and degree of implementation among classrooms within any given program. This analysis would be facilitated by the existence of a uniform instrument used for all classrooms part of the same program. Among programs, however, this relationship would be more difficult to assess to the extent that the implementation checklists and other assessment instruments vary from program to program. To the extent that such among-program degrees of implementation assessments can be made, this factor will also be analyzed. All classrooms studied will be part of the analysis conducted for this particular issue.

Data Reduction

Data obtained throughout the 1976-77 school year from pre-test administration and from periodic observation of program operation and characteristics will be reduced for analysis upon collection. Specifically, data obtained from pre-test, initial interviews, and the initial observation cycles at all sites will be processed in a manner that will allow the data reduction and analysis procedures to be evaluated early during the study year. Procedures for the speedy reduction of the post-test results and other late school year data will be carefully refined during this school year to ensure that once the data are obtained in May 1977 all necessary data are reduced and ready for analysis in order to report the study's results in July 1977. Since TURNKEY experience has indicated the extremely critical nature of these data reduction tasks in the past, sufficient resources will be allocated to this activity in our proposed design to ensure their timely completion.

Once all data has been obtained, the carefully planned data analysis effort will proceed. The practical details of this critical element of this study are addressed further in Section IV of this report. Appendix 13 includes examples of the data collection/coding/reduction forms modified from forms used by TURNKEY elsewhere.

Implications and Limitations of the Approach

The nature of the sampling plan, especially its reflection of our Definitional Matrix, allow a number of issues critical to NIE and the prime audience of the proposed study (e.g., Congress) to be addressed in a highly structured manner. Though a true experiment wherein all else was kept constant and only one program feature was systematically varied at any given time and students were assigned to treatments randomly is impossible to achieve in any real world setting, the quasi-factorial structure suggested here allows an approximation of at least a portion of this scheme to obtain. (The reason for not fully utilizing the factorial-interaction potential of our Definitional Matrix was discussed in detail earlier.) Careful selection of programs and matching of students in adherence to the sampling constraints listed earlier will have to suffice as the suggested approximation of random assignment of students, recognizing that this approximation has been a major problem for all large-scale field research and evaluation studies in education.

The analyses and contrasts delineated above should provide adequate answers for NIE to meet its obligation to Congress and for Congress to effectively tackle the task of writing the new comp ed legislation in 1977. Other contrasts are clearly available as well given the proposed design framework. For instance, the issue of whether the processes included in our definition of individualization have different effects in mainstream vs. pull-out programs could clearly be investigated by combining the logical framework of Level 3 with that of Level 2. Only the contrasts considered most basic have been specifically listed here, all other possibilities remain for the research team to consider throughout the implementation phase.

References

Section III

Averch, H.A., Carrol, S. J., Donaldson, T. S., Riesling, H. J., and Peneno, J. How Effective is Schooling? A Critical Review and Synthesis of Research Findings. Santa Monica: Rand, 1971.

Bloom, B. S. Stability and Change in Human Characteristics. New York: Wiley, 1964.

Bond, G. L. and Dykstra, R. Final Report: Coordinating Center for First Grade Reading Instruction Programs. Minneapolis: University of Minnesota, 1967.

Brookover, B. B., Thomas, S., and Patterson, A. Self-Concept of Ability and School Achievement. Sociology of Education, 1964, 37, 271-278.

Carroll, J. B. A Model for School Learning. Teachers College Record, 64, 1963, 723-733.

Dryer A. S., Haupt, D. Self-evaulation in young children. Journal of Genetic Psychology. 1966, 108, 185-197.

Holt, J. How Children Fail, New York: Delta Books, 1964.

Kelley, Edgar A., Wood, Fred H., and Joekel, Ronald. Teacher Perceptions of School Climate and the Implementation of Individually Guided Education (IGE). Unpublished doctoral dissertation. University of Nebraska, 1973.

Klein, Stephen. Procedures for Comparing Instructional Programs. Los Angeles: UCLA Center for the Study of Evaluation, 1972.

McDaniel, E.D., et a. Longitudinal Study of Elementary School Effects. West Lafayette, Indiana: Purdue University, 1973.

Neuman, A.P. "Later Achievement Study of Pupils Under-Achieving in Reading in First Grade." Reading Research Quarterly. 1972, 7, 477-508.

Popham, W. J. "Program Fair Evaluation -- Summation of Assessment of Instructional Sequences with Dissimilar Objectives." National Society for Programmed Instruction, 1969, 6, 1-9.

Wylie, Ruth C. "The Self Concept; a Critical Survey of Pertinent Research," Liternature, Lincoln, Nebraska: University of Nebraska Press, 1961.

SECTION IV

LOGISTICAL ISSUES/TASK APPROACHES

In Section III, we have specified the theoretical and analytical framework within which the implementation of the individualized instruction study will be conducted. In this section, we describe the operational aspects of the study and its interaction with the participating programs. Specifically, we address the data collection procedures to be used (Task 6), techniques for enlisting site cooperation (Task 7), implementation project staff (Task 8), and the overall project schedule (Task 9).

TASK 6 -- DEVELOP DATA COLLECTION PLAN

In Tasks 1 through 4, discussed in Section II of this report, we have performed the first preliminary data collection step -- the identification of the study's data (and hence data collection) requirements. As indicated in our Proposal, these needs fall into three principal categories:

- Program and School Characteristics -- management and operational factors
- Program Implementation Information -- level-of-implementation data on the degree to which the study programs are operating as intended
- Program Effectiveness Measures -- program outcome measures of student performance and attitudes

Also in the first four tasks, we discussed the detailed factors involved in the development of the specific instruments to be used to gather the needed data. In Section III/Task 5, we specified the procedures to be used to analyze the data, once it has been collected. In this task, we outlined the data collection plan -- the logistical procedures which bridge the gap between the data collection instruments and the analysis plan.

The data collection plan for this study of individualized instruction has five major components:

1) Selection of Field Staff
2) Training of Field Staff
3) Data Collection Procedures
4) Field Cooperation
5) Data Confidentiality

The first two of these components relate specifically to the staffing of the project and will consequently be discussed in Task 8, the development of a staffing plan. The fourth component, gaining the cooperation of school district personnel at each site, will be discussed as an individual topic in Task 7. In this task we will, therefore, discuss only the data collection techniques to be used and the procedures to be followed in ensuring data confidentality.

Data Collection Procedures

As indicated in Section III/Tasks 1-4, data collection activities will consist of three general instruments/techniques:

- staff interviews
- testing and affective measurement
- classroom observation

The data collection procedures will reflect the nature of the data collection instruments/techniques to be used. After preliminary subtasks, in which the instruments are finalized, other preparation activities conducted, and the instruments/techniques field tested, interviews of school district staff members will be conducted, pre- and post-tests of students will be administered, and classroom observations will be performed. Each of these specific subtasks will be described in turn. It should be noted that it is possible to specify the procedures to be used although some of the instruments themselves may not be finalized. Some of the procedures described herein will be of a general nature, to be finalized when the first data collection subtask, the finalization of all data collection instruments is conducted as indicated in Section III/Task 3.

Finalization of Data Collection Instruments

Since timing is a critical factor in the performance of this study, it is important that the instruments to be used in the collection of data be finalized as early as possible so that the necessary and time-consuming OMB clearance procedures may be completed without delaying the project. For all intents and purposes, it is possible to divide the recommended instruments (see Section III/Tasks 1-4) into two categories: those which may be finalized before site selection with very little work on the part of the implementation contractor, and those which are site-dependent and which must consequently await final site selection.

Most of the instruments fall into the first of these categories. Achievement tests and specific affective measures (discussed in Section III/Task 4) are all to be used in their existing forms and, therefore, will require no modification (some modification in traditional scoring and analyses procedures may be needed, but these would not affect the form or administration of these existing instruments). The interview guides for interviews of principals, teachers, aides, and other staff members have been specified to some degree of detail in Section III/Tasks 1-4 (an example of a finalized interview instrument is shown in Appendix 13) and will require only minimum modification by the implementation contractor. The modification of these instruments should take no more than eight weeks after implementation contract award and will, therefore, be ready for submission to OMB for approval and field testing in selected school districts during April 1976.

The second category of instruments, those whose final design is dependent upon the sites included in the final sample, consists only of the implementation instruments which are to form the basis for inputs into interview and classroom observation instruments (see Section III/Task 3). It is our belief that only be investigating a program's operational characteristics as they relate to the program's operational intent can an evaluator truly assess the program's level of implementation. The diverse nature of the individualized programs, which are expected to be included in the study, indicate that it is highly desirable for these instruments to be designed specifically for each program. Since the final selection of the sample districts, schools, and classrooms is estimated to be completed in May 1976, (see Task 9) the instruments themselves cannot be fully developed until the Spring of 1976. This is not the liability it might seem, however. These instruments/checklists will not require OMB clearance, since the completion of these instruments does not require specific interaction with classroom staff members; hence it may be viewed as merely data collection notations rather than true instruments. Also, since each of these classroom observation instruments will be unique to its specific program and/or LEA it may not fall within OMB's requirement that clearance is needed for instruments being used in ten or more LEA's. As a alternative to this approach of developing instruments specific to each program, we have (as is indicated in Section III/Task 3) developed a program-independent instrument.

For all data collection instruments which require OMB clearance, the implementation contractor will prepare a complete OMB documentation package. It should be noted that should NIE wish to do so, it could approve the suggested data collection instruments and submit them to OMB prior to the selection of the implementation contractor. We believe, however, that it is desirable for the implementation contractor to have some input into the instruments to be used. This delay until Spring 1976 for OMB submission will not delay the project in any way since it does not fall on the study's critical path. All instruments will be ready for use when school begins in September 1976.

Field Test of Data Collection Instruments

During the Spring of 1976, the data collection instruments and procedures, will be field tested in two school districts.

These field tests will be limited to those instruments which have been modified or developed for this study. The student outcome instruments, the achievement and affective measures need not be included in the field test for two reasons: 1) the administration of achievement and affective tests to students at field test sites might unduly burden the classroom staff, and; 2) each of the suggested student outcome instruments has already been field tested and validated many times before.

Specifically to be included in the field test are the interview instruments for principals, teachers, aides, and other staff members, and the classroom observation instruments. The field test of the classroom observation instruments implicitly denotes a field test of the procedures by which the classroom observation instrument/implementation checklist is determined for each program (see Section III/Task 3).

It is suggested that one individualized and one standardized program be selected in each of the two field test districts. One of the districts should have two reading programs, while the other should have two mathematics programs. Within each of these programs, the recommended field test will include interviews with the school principal, at least two teachers (including at least one regular classroom teacher who serves compensatory education students plus at least one specially hired compensatory education teacher), and at least one aide (if applicable), as well as of the district-level Federal program director. Also within each of the two programs at each field test site, at least three classroom observations will be conducted.

Sites for the field tests should be drawn from the list of programs suggested for inclusion in the study. This will ensure that the field test sites not only have an individualized instruction program, but that enough information will be available for the development of classroom observation instruments/implementation checklists for each program. For purposes of economy and usefulness, the two field test sites should be relatively close to the implementation contractor's major office and should differ somewhat in the general nature of their population.

The logistics of the field test indicate that plans for its conduct should be undertaken immediately after implementation contract award. A member of the implementation contractor's project team should be designated as Field Test Director. This Field Test Director will be responsible for contacting each of the field test sites and arranging for their cooperation in the field test. This field test director will serve as continuing liaison between the field test sites and the implementation contractor, and may even administer the field tests of specific instruments personally.

It is expected that three principal benefits will emerge from the field test. The first is the obvious benefit of being able to provide operational support for the OMB documentation package. The second, less obvious, benefit is the insight gained into the procedures associated with the administration of many of the data collection instruments. It is possible to capitalize on this insight by making the Field Test Director an integral part of the project group which plans and conducts the training sessions for data collectors. A third, and even less obvious, benefit of the field test should be an understanding of the incentives which might be needed to ensure the willing participation and cooperation of district, school, and classroom staff.

Preparation for Data Collection

An often overlooked but extremely important aspect of any major data collection effort is the laying of groundwork for proper communication and coordination throughout the project. The purpose of this subtask is to keep appropriate groups and individuals abreast of the project and to ensure that scheduling of project activities is performed as expeditiously as possible.

It is recommended that before specific contact is made with project sites, the implementation contractor contact the Committee on Evaluation and Information Systems (CEIS) of the Council of Chief State School Officers

(CCSSO). Not only will this initial contact satisfy protocol, but it is quite possible that SEA staff may be able to suggest additional well-implemented individualized programs to be considered as project sites.

The first contact with each of the districts after their selection as project sites should be by telephone using the LEA official designated in the sampling questionnaire as the appropriate contact. In most areas, this contact person will be the district's Director of Federal Programs. This initial site contact should take place immediately upon the selection of each site for inclusion in the final sample, during April and May 1976. The purpose of this initial telephone contact is:

- to inform that district of its selection and confirm its participation desire;
- to formalize the communication channel between the district designee and the Data Collection Manager of the implementation contractor;
- to specify the schools and classrooms which have been chosen as study units;
- to offer to assist in explaining the project to local groups (e.g., parents), through on-site presentation;
- to negotiate the specific incentive arrangements to be used to facilitate staff cooperation (see Task 7); and
- to solicit district input into the selection of on-site data collection staff.

It is believed that by enlisting the district's designee as a partner in the data collection effort, the implementation contractor will minimize the probability of major disruptions during the course of the project.

After contacting the district-level designee, the implementation contractor should make initial contacts with the principals of the schools in which data is to be collected. This contact might best be made in two steps: first, an initial telephone contact by the implementation contractor's Data Collection Manager (see Task 8), informing the principal of his school's selection and introducing the local data collector; and, second, a brief visit to the school by the local data collector, enabling him/her to become familiar with the school, to establish open communication with the principal, and to tentatively schedule time for the conduct of initial staff interviews and the administration of pretests for achievement and affective instruments.

It is expected that a second preliminary school visit will be made by the local data collector during the district's teacher orientation week (generally the last week in August). This visit will enable the local data collector to introduce him/herself to the teachers whose classes are involved in the study, to explain the overall nature of the study, to respond to any concerns the teachers may have about classroom disruptions, and to confirm the schedule for interviews and test administration. This visit may involve meeting with teachers either individually or as a group, depending on the preference of the local data collector, the principal, and the teachers themselves.

Although it has not been listed here as a specific preparatory activity, budgetary provision should be made for at least one additional (beyond the sampling visits) visit by a member of the implementation contractor's core project team. Experience tells us that such a visit can do much to facilitate the liaison between the local data collector and site personnel. Although such a visit may not be necessary in all cases, it will frequently serve as healthy reinformcement for the initial training session attended by the local data collectors.

As a final preparatory step, arrangements for computer facilities will be made. The importance of the time factor makes it imperative that rapid data turnaround be available. For this reason the DHEW computer will not be used. A primary and backup facility will be established to provide insurance against last minute logistical problems.

Conduct of Staff Interviews

The first data collection activity to take place is the interviewing of school and classroom staff. These interviews should, if possible, be conducted during the first four weeks of the 1976-77 school year. Indeed, it may be possible to conduct some or all of these interviews in conjunction with the introductory school visits just described. It is important that these interviews be used not only to gather data, but also to establish a cordial relationship with teachers, which will facilitate the later administration of tests and conduct of classroom observations.

In order to reflect any changes in attitudes or perceptions on the parts of school personnel, the same or parallel interviews will also be conducted at the end of the 1976-77 school year.

Although the specific procedures to be followed in interviewing school personnel will be different at each site, some basic precepts, gained from years of such interview experience, may be stated. Whenever possible, times for interviews should be scheduled at least two days (and preferably a week in advance). Although principals generally have the flexibility to be available when needed for interviews, they are frequently interrupted by day-to-day school problems; it is generally advisable to allow extra time (at least 60 minuites) for such principal interviews. Teachers are generally available to be interviewed during their free period if it is scheduled in advance. The teacher's lounge or an empty classroom is frequently the most suitable place for such interviews. Since aides have less free (or planning) time than teachers, it is generally more difficult to interview them. If properly scheduled, however, it is possible for the teacher to arrange the aide's time in such a way as to permit an interview during the regular school day.

In terms of the time required of the local data collector for staff interviews, the following estimates have been made:

- Approximately 40 classrooms in each district will be included in the study sample.
- Approximately 10 principals, 60 teachers, and 20 aides will be interviewed in each LEA.
- About five interviews can be conducted each day (allowing for difficulties in scheduling interviewees).
- The series of interviews will be conducted once at the beginning and once at the end of 1976-77 school year.

The net result of these estimates indicates that between 35 and 40 days will be spent by the local data collector interviewing school staff. Experience also indicates that an additional 15 days will be needed to code the resulting data and convert into a form suitable for input into the analysis plan. In order to reinforce the training sessions, it is planned that the cognizant Regional Coordinator will provide on-site assistance to the local data collector in the conduct of early interviews. This type of monitoring will also serve to enhance the cross-site reliability of interviews.

Administration of Tests

The second data collection activity is the administration of the achievement and affective test instruments specified in Section III/Task 4. In order that the results of the study maintain credibility, it is critical that the test results avoid contamination by inadequate testing procedures. Since all testing instruments are already developed, test administration guides also exist for all of them. These test administration guides will be discussed between the local data collector and each teacher (This discussion may take place during the interview.), and the appropriate schedule for test administration will be developed. In order to avoid variations in the time between pre- and post-tests, it is desirable that corresponding tests be administered at approximately the same time -- all within a one-to-three-day period -- following the suggested time periods in the manuals which are found on the "norming" test schedule.

It is estimated that approximately ten working days will be required of the local data collector for the scheduling and administration of the pre-test, with a like time require for the post-test. In order to conduct the testing in the manner described, the Regional Coordinators will assist the local data collectors and the teachers during the pre- and post-test phases of data collection.

Conduct of Classroom Observation

Frequent observations of all study classroom will be conducted during the months of September 1976 through April 1977. (These classroom observations will be conducted on an irregular basis but scheduled approximately two days prior to the visit.)

The principal purpose of the classroom observations is to assess the degree to which the programs are implemented according to their plan and to verify certain time schedules provided by teachers during these interviews. The specific instruments to be used in recording the classroom observations will, in part, be uniquely determined for each program, based upon the specific goals, objectives, and procedures identified for the program. These instruments have been discussed in Section III/Task 3.

In terms of local data collector time required for classroom observations, we have made the following approximations:

- All observations will be for an entire period.
- There will be about 40 classrooms in each district participating in the study.
- All classrooms will be observed four times from September 1976 to April 1977.
- No observation of a classroom will occur within 30 days of the previous observation of that classroom.
- An attempt will be made to conduct the first classroom observation as early as possible after the program stabilizes -- generally around the third week in September.
- It is possible to conduct approximately four full-period classroom observations per day.

These approximations indicate that about 40 days during the project will be spent by each local data collector in classroom observation. Experience also indicates that an additional ten days will be needed for coding and other administrative tasks directly associated with the classroom observations. Again to ensure observation reliability and to reinforce the training sessions, the Regional Coordinators will team with the local data collectors during the first round of observations at each site.

Data Confidentiality

Confidentiality of educational data -- particularly student-related data -- is a subject which has gotten a great deal of attention during the past year. In order to address this issue, we feel that an on-site coding system is the best insurance against improper use of data.

At each site, the local data collector will, during the first week of the 1976-77 school year, develop a keyed document which will assign a unique code number of each school, principal, teacher, aide, and student involved in the study. This key code containing the unique identifiers for all study participants will be kept in a secured area in the local data collector's office. Only one additional copy of the key code will be made. This copy will be sent in a clearly labeled sealed envelope to the Project Director. The Project Director will maintain the key codes from each site (unopened) in a secured area, as a precaution against inadvertant destruction of an on-site key code. Immediately upon termination of the project, both copies of all key codes will be destroyed.

All data collected on-site will be coded with the unique identifier(s) of the study participants involved. No data forwarded to the project core staff will contain names of individuals -- only coded identifiers.

Administrative Procedures

In addition to the specific data collection instruments identified in Section III/Tasks 1-4, other administrative reporting procedures will be needed and will be the responsibility of the local data collector. There will be three administrative reports which will be maintained:

- Time Allocation Log
- Expense Report
- Daily Activity Log

The Time Allocation Log, shown in Exhibit IV-1, is a management control mechanism by which on-site project activity can be monitored. It also serves as a time sheet input into the implementation contractor's financial system for the local data collector. Each of the activities which the local data collector will be expected to perform is enumerated (along with miscellaneous categories designed to include coding and other related activities). The Expense Report will be the standard expense form for the implementation contractor, including the nature and amount of each reimbursable expenditure. The Daily Activity Log, shown in Exhibit IV-2, allows the local data collector to maintain a record of his/her individual data collection contacts. This log will be maintained in coded form (i.e., all schools and staff members will be identified only by their coded identifiers). Each of these administrative reports will be completed and sent to the local data collector's assigned Regional Coordinator (see Task 8) on a bi-weekly basis.

Exhibit IV-1

TIME ALLOCATION LOG

name ______________________

district ______________________

period ______________

ACTIVITY	1/16	2/17	3/18	4/19	5/20	6/21	7/22	8/23	9/24	10/25	11/26	12/27	13/28	14/29	15/30	31	TOTAL HOURS
INTERVIEWING																	
INTERVIEWING-MISC.																	
TESTING																	
TESTING-MISC.																	
OBSERVATION																	
OBSERVATION-MISC.																	
TRAINING																	
ADMINISTRATIVE																	
OTHER (specify)____ ______________																	
TOTAL HOURS																	

Exhibit IV-2

DAILY ACTIVITY LOG

name ______________________________ **date** ______________________

district ______________________________

---

1 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

2 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

3 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

4 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

5 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

6 TYPE OF ACTIVITY: Interview Testing Observation

Location: ____________________ (School) ____________________ (Staff)

Comments: ____

All School and Staff enteries are to be denoted by numerical identifiers only.

Each of the administrative instruments/procedures described in this data collection plan will be discussed in somewhat greater detail in the Data Collector's Manual described in Task 8.

TASK 7 -- PROVIDE LOGISTICAL PLAN FOR PROGRAM COOPERATION

In this task, we discuss several problems associated with obtaining cooperation from study participants and outline a set of procedures and incentives designed to ensure this site cooperation. These perceptions are based upon the experience of project team members who have worked with more than 200 LEAs during the last five years and upon related experiences in several similar large-scale Federally sponsored studies. In describing suggested procedures, frequent reference will be made to other tasks, in which the procedures are described in detail.

General Impediments to Cooperation

Many of the LEAs and publishers contacted during the design phase have been somewhat concerned about site participation in the proposed study. Several LEAs reflected upon "bad experiences" in previous national studies; some expressed concern over the role of NIE in evaluating programs which are in large part locally-developed and locally funded; still others were somewhat concerned about outsiders coming into their district. Several firms and publishers also expressed the following types of concerns:

- Why is there a need to evaluate our program again? Independent evaluations have already proven its success;
- Will the contractor involve our consultants or Area/Regional manager, before going on-site, in order to "understand" the program thoroughly?;
- May we review instruments used to assess program success prior to final selection to ensure that the contractor is measuring what the program is designed to accomplish?

While the reasons for the concerns may or may not be valid, these varying perceptions must be taken into account.

A second concern relates to the potential disruptive nature of the study, especially if a large amount of observation is to be conducted. The requirements to be placed upon the general administrative staff and principals (especially related to any data collection from parents) appear to be of primary concern.

A last general concern appears to stem from new legislation, guidelines, and judicial decisions related to confidentiality of data. In any new area where uncertainty of interpretation exists and individuals may be held personally liable, anxieties will surface. Several LEAs said they would not participate unless it were made impossible to identify individual respondees; others, wanted their legal advisors to review any district data prior to public release. The implementation contractor must address these issues and demonstrate strict adherence to "contractor guidelines" now being prepared by HEW.

Specific Concerns

Based upon our discussions with LEA staff members and consultants (including a research-oriented Principal of a Title I school, a Federal program director in a middle-size city and program development specialists in an inner-city district) during the site identification/initial selection, a number of concerns related to either individualized or standardized programs were highlighted.

Individualized ("I") programs are more likely to be accustomed to outside visitors and evaluation, especially if they have been selected for review by the USOE DRP or PIP programs. On the other hand, too great a study requirement on an already over-burdened school staff could erode cooperation. Conversely, standardized ("S") program staffs might be very anxious and uncooperative, especially if they perceive their role as that of a control or comparison school. "I" program staff are likely to be more inquisitive (e.g., "How does our program compare to other similar programs in the study?") and to seek constructive criticism. Care must be taken to minimize the influence of outsiders and unintended effects generated during the period of observation.

Since testing is generally more of an integral component of "I" programs, concern over additional testing may be lower. The implementation contractor must be able to demonstrate a clear need to administer the additional instruments, or the LEA staff may perceive the testing as being merely disruptive. Staff in "S" programs are likely to be more suspicious of testing beyond minimal LEA requirements and more concerned about the improper use of test scores. To ensure LEA cooperation, the administration of any additional tests must be clearly justified to LEA staff; the use of resulting data must be limited to the study.

Procedures for Enlisting Cooperation

During the design phase, several measures were taken to increase the probability of cooperation and good working relationships with those LEAs finally selected for participation. Data requests were limited to only those necessary to meet the purposes of initial site selection. Although time-consuming, a review of existing documentation of projects was used to complete the checklists. Thus, when telephone contacts were made, the questions asked were limited to verification of existing data or a small number of direct inquiries, thereby reducing the length of conversations. As an example, the final checklist completion for the three programs in one district took eight minutes, while review of documentation took nearly a half day. All LEA contacts were open, frank, and unofficial. Several LEA staff members wanted to get approval from their Board of Education prior to releasing any information. We emphasized the tentative nature of their potential participation, indicating the steps which are being followed, including a formal information request from NIE or the implementation contractor next year. We suggest that the sampling variables and general conditions for participation be determined as quickly as possible by NIE and that survey instruments similar to those in Appendix 5 be cleared by OMB and sent to potential sites as soon as possible. One person at NIE or the implementation contractor should be available by telephone to answer questions regarding the data request and possible district participation. Since many of the candidate programs are at sites participating in other Federal efforts, this coordination at the Federal level will also tend to minimize confusion.

The primary means of enlisting the support and cooperation of LEA staff is through careful selection of local data collectors who will represent the primary liaison between the district and the project. The specific procedures and qualifications of local data collectors who will represent the principal liaison between the district and the project. The specific procedures and qualifications of these field staff will be discussed in Task 8. It will suffice in this task to indicate that among the qualifications for the local data collectors will be the ability to work closely and cordially with LEA staff at the district, school, and classroom level. It is strongly felt that the use of local data collectors, familiar with the policies and traditions of the district as well as, in many cases, its personnel, is a major step in avoiding the "outside evaluator" stigma frequently associated with such field studies.

At least two types of incentives for cooperation may be provided to LEA personnel. The first type of incentive is informational in nature: the provision of informational feedback to the LEA on the performance of their programs. Two major concerns must be exercised when considering such an incentive. The first is that NIE is the primary consumer of the study's results and that any ancillary use of study data or results must meet with NIE's full approval. The second concern is that any early feedback of data or results may have a confounding effect on the study by influencing the subsequent behavior of study participants. It is our recommendation that, with NIE approval, project results in draft form be provided to the participating LEA staff as soon as possible after the posttest is administered.

The second type of incentive is monetary in nature. It would provide a nominal sum to participants in gratitude for their cooperation. A number of alternative means of providing this incentive present themselves. One technique, frequently used, is simply paying teachers a small sum (perhaps $10) for submitting to an interview. Although this does distribute funds to teachers, it is unlikely that, on an individual basis, this small sum would provide any great incentive for project cooperation. Another means of providing incentives for staff cooperation, and one which we recommend, is to provide the incentive on a program-wide basis, distributing a total of perhaps $250 per program to be allocated equally among the participating school buildings. This would provide the funds as an addition to the regular resources available to the students participating in this. Care must be exercised, when distributing this money, not to disrupt the LEA's normal financial system. Close coordination with the superintendent's district-level designee will be observed. A second concern is that the introduction of such dollar incentives might influence the conduct of instruction by introducing additional resources into the classroom. Although this is unlikely, the possibility may be addressed by not distributing the incentive funds until after the study is complete.

TASK 8 -- DEVELOP STAFFING PLAN

In this task, we specify the staffing plan which will be required to conduct the implementation of the study of individualized instruction in the manner described in this report. Our discussion of this task can be broken down into five principal subareas:

- Project Organization
- Project Person Loading
- Selection of Field Staff
- Training of Field Staff
- Data Collector's Manual

As indicated in our discussion of Task 6, the last three of these subareas might also be considered part of the data collection plan.

Project Organization

Exhibit IV-3 depicts a suggested organization chart for the successful implementation of the study. This activity-oriented structure reflects the three major tasks to be performed in the study's implementation phase (i.e., Sampling, Data Collection, and Analysis) as second level management.

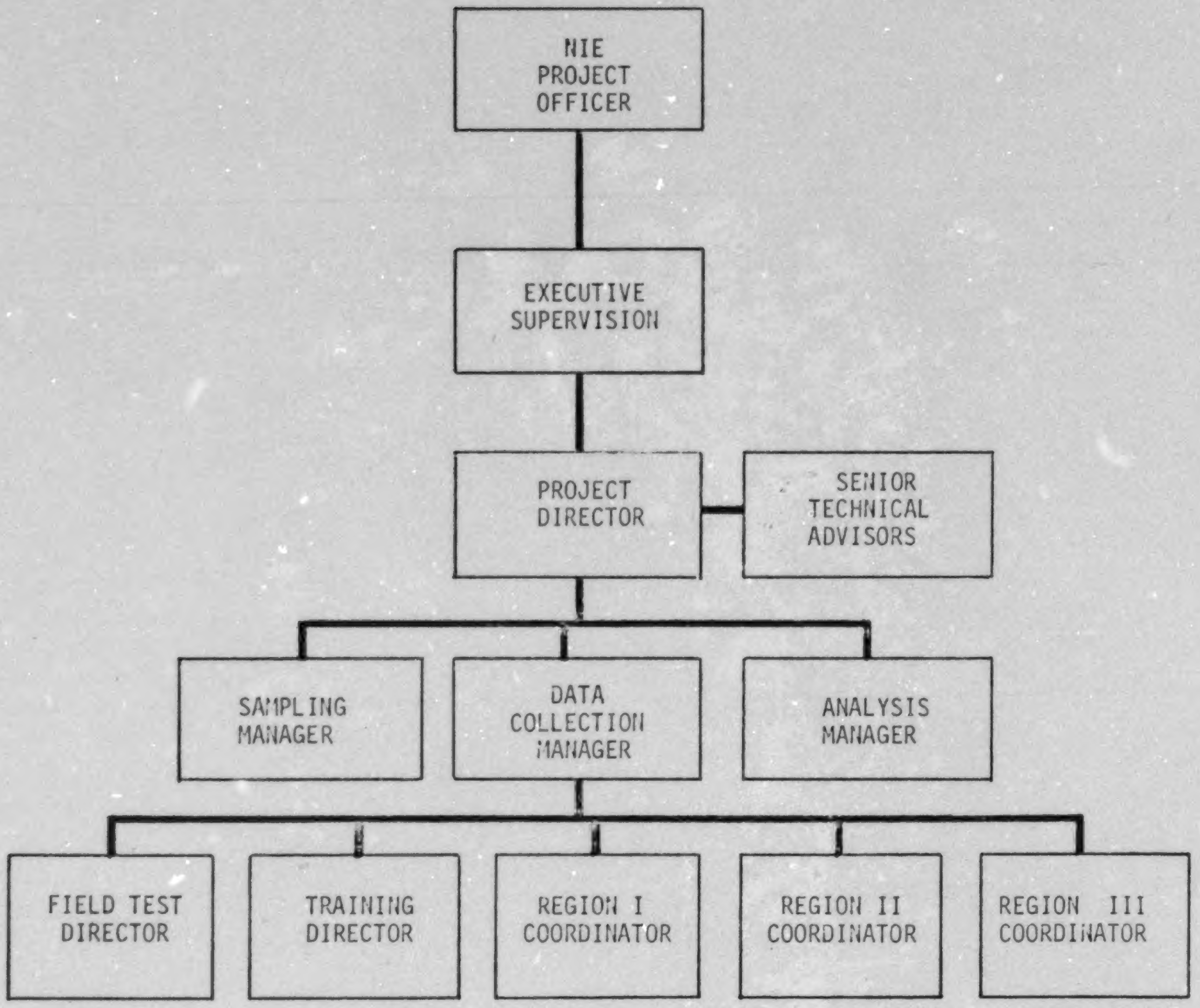
The nature of the position and the necessary qualifications of each member of the core staff will be discussed in the following paragraphs:

Executive Supervision

As with any large-scale project, it is desirable to have top level oversight from someone at a high level within the management of the con-

EXHIBIT IV-3

ORGANIZATION STRUCTURE


NIE PROJECT OFFICER
EXECUTIVE SUPERVISION
PROJECT DIRECTOR
SENIOR TECHNICAL ADVISORS
SAMPLING MANAGER
DATA COLLECTION MANAGER
ANALYSIS MANAGER
FIELD TEST DIRECTOR
TRAINING DIRECTOR
REGION I COORDINATOR
REGION II COORDINATOR
REGION III COORDINATOR

tractor. The role of the person providing this Executive Supervision would be:

- to make official commitments on the part of the implementation contractor,
- to serve as liaison between the contractor and the NIE Contract Officer in contractual matters,
- to commit the appropriate staff of the implementation contractor to the study, and
- to serve in an advisory capacity on technical matters to the Project Director.

Executive Supervision of the study would have little or no technical involvement and would participate at a relatively low level of effort (perhaps three days per month). The person filling the Executive Supervision role would have such a title as Project Administrator, Partner-in-Charge, Vice-President, or even President within the implementation contractor's management organization.

Project Director

The Project Director will be the person directly responsible for the technical performance of the study. The principal functions of the Project Director will be:

- to serve as the major point of contact between the NIE Project Officer and the implementation contractor's project team,
- to allocate the personnel and material resources of the project team to best serve the successful implementation of the study,
- to provide technical leadership in the design, development, and conduct of project tasks,
- to monitor the day-to-day progress of the study, and

- to serve as principal investigator/author of the final evaluation report.

The Project Director should be experienced in both compensatory education and the evaluation of education programs. Even more important,he/she should be experienced in the management of similar large-scale evaluation studies. In the interests of project quality and credibility in the academic community, it would be desirable if the Project Director held a doctoral degree or five years experience in directing similar large-scale technical studies. The Project Director should devote at least 75% of his/her time to the project for the duration of the study.

Senior Technical Advisory Panel

In a project such as this study of individualized instruction, it is extremely desirable for the Project Director to have at his disposal a number of well-known experts in the various disciplines associated with the study, as well as others whose interests are closely associated with the study. These advisors would assist the project team by providing special expertise when needed and general project review and criticism.

This advisory panel would have special expertise in such areas as individualized learning, compensary education, tests, measurement, and evaluation. Its members could also include persons chosen for their insight into the political as well as technical aspects of the project. A suggested size for this panel would be approximately ten people who would devote their efforts to specific tasks designated by the Project Director.

Sampling Manager

The first task in the implementation phase of the study (see Task 9) is the selection of the final sample from among a nominated group based upon existing program documentation and sampling data gathered during the early stages of the implementation phase. The Sampling Manager should have experience in sampling theory and techniques, as well as in design and analysis of projects involving LEA's. The Sampling Manager should hold at least a master's degree in education, evaluation, quantitive methods, or some related discipline. The Sampling Manager will have heavy involvement during the early stages of the project with very little work after the sites have been chosen. The nature of the functions of the Sampling Manager and the Analysis Manager make it possible for one person to fill both roles.

Data Collection Manager

The Data Collection Manager will have responsibility for the conduct of the second and third tasks (see Task 9) of the implementation phase of the study, including the selection and training of field staff and the management of the overall data collection effort. The Data Collection Manager should be experienced in survey research and data collection procedures and in the logistical aspects of interacting with LEAs. The Data Collection Manager should hold at least a master's degree in education, survey research, or some related field. The Data Collection Manager will essentially devote full time to the project for its duration.

Analysis Manager

The Analysis Manager will be responsible for the fourth task of the implementation phase, data analysis. The Analysis Manager should be experienced in the evaluation of education programs in general and in the testing and measurement of educational achievement and attitudes in particular. The Analysis Manager should hold at least a master's degree in one of those disciplines. The involvement of the Analysis Manager will be fairly uniform during the course of the project, with particularly heavy involvement during the finalization of data collection instruments and the reduction and analysis of actual data. As indicated, the Analysis Manager might, if qualified, also fill the role of Sampling Manager.

Field Test Director

The field test of data collection instruments is to be conducted during the Spring of 1976. One member of the project team, with experience in the various data collection aspects of public education, should be designated as director of that field test. He/she would then have responsibility for all activities involved in the field test and for ensuring that field test results are fed back to the appropriate members of the project team, as well as coordinating the field test results with the OMB documentation package. The Field Test Director will have heavy project involvement during the first four months of the implementation project. The position becomes inoperative when the field test is completed and the results reported. The Field Test Director may also be qualified to fill the role of one of the Regional Coordinators.

Training Director

A crucial facet of the data collection effort is the training of field staff. Local data collectors will be trained in two formal training sessions which must be properly planned and executed. One member of the project team with expertise in this area should be specified as Training Director. The Training Director will devote most of his effort during the Summer of 1976.

Regional Coordinators

In order to effectively monitor the large magnitude of data which is expected to flow from the sites and to ensure that the highest possible degree of data reliability is achieved in the data collection efforts, three Regional Coordinators should be assigned from the implementation contractor's staff. These Regional Coordinators will work closely with the local data collectors in their regions. They may also team with the local data collectors for the actual collection of some data. Each Regional Coordinator will have responsibility for monitoring and assisting approximately eight sites. The exact assignment of sites and programs will, of course, await final site selection. These Regional Coordinators will devote essentially full time to the project during the data collection period (i.e., the 1976-77 school year).

Project Person Loading

Based upon our assessment of the project requirements, the suggested scope of the analysis, and the estimated size of the sample, we have made a task-by-task estimate of the level of effort required by each project team member. Exhibit IV-4, following this page, reflects these person-loading estimates.

Selection of Field Staff

A critical component of the overall project staffing is the selection of properly experienced and qualified field staff as local data collectors. As discussed in Task 7, we believe the use of professionals with community ties and acquaintances in the school districts chosen as study sites will greatly enhance the level of cooperation received from site personnel. It is important, however, that these field staff be more than merely acquainted with the schools and its personnel. They should have familiarity with compensatory education programs and should be aware of the political and administrative implications of education evaluations such as this one.

Since it is not possible at this time to specify with certainty the number of study units (classrooms) included at each site we have made estimates of time requirements based upon approximately 40 classrooms per site Based upon the data collection activities noted in Exhibit IV-4, this estimate includes the following field staff requirements during 1976-77 school year for an average site.

Exhibit IV-4

ESTIMATED PROJECT PERSON LOADING

tasks	person days							
	EXECUTIVE SUPERVISION	ADVISORY PANEL	PROJECT DIRECTOR	SAMPLING** MANAGER	DATA COLLECTION MANAGER	ANALYSIS** MANAGER	STAFF***	LOCAL DATA COLLECTORS
TASK 1: SELECT FINAL SAMPLE								
1.1 Administer Questionnaire	2	--	7	25	5	5	40	--
1.2 Conduct Verification Visits	1	--	9	20	5	--	40	--
1.3 Select Sample	4	20	16	20	5	--	10	--
TASK 2: SELECT AND TRAIN DATA COLLECTORS								
2.1 Interview Local Candidates	--	--	4	--	5	5	40	--
2.2 Select Data Collectors	2	--	4	--	10	5	10	--
2.3 Develop Training Materials	6	20	40	--	30	20	180	--
2.4 Conduct Training Session I	1	--	10	5	10	5	20	125
2.5 Conduct Training Session II	1	--	10	5	10	5	20	--
TASK 3: CONDUCT DATA COLLECTION								
3.1 Finalize Instruments	6	--	20	--	20	20	85	--
3.2 Prepare for Data Collection	2	--	8	--	35	5	60	875
3.3 Conduct Interviews	1	--	20	--	35	10	320	1375
3.4 Conduct Testing	2	--	20	--	30	10	180	500
3.5 Conduct Observations	3	--	70	--	70	10	400	1250
TASK 4: CONDUCT DATA ANALYSIS								
4.1 Reduce Data	--	--	15	--	15	25	60	--
4.2 Obtain Scored Test Results	1	--	5	--	5	5	20	--
4.3 Check Data for Consistency	1	--	10	--	5	10	140	--
4.4 Conduct Specific Analyses	8	20	15	--	5	20	180	--
TASK 5: PREPARE FINAL REPORT								
5.1 Prepare Draft Final Report	6	--	10	10	10	10	30	--
5.2 Prepare Final Report	3	20	25	10	20	10	20	--
total	50	80	310	95	330	220	1865	4250
percent of time*	14%	--	86%	26%	92%	61%	--	--
total project person days								7200

* Based on total project duration of 360 work days

** If a qualified person is available, the roles of Sampling Manager and Analysis Manager may be filled by the same person.

*** Including Regional Coordinators

	Person-Days	
Interviewing	40	
Interviewing-Misc.	15	
Testing	10	
Testing-Misc.	10	
Observation	40	
Observation-Misc.	10	
Training	10	
Administrative	35	
	170	TOTAL

These estimates represent average time requirements for the local data collector at each site to conduct each of the three primary data collection activities (i.e., Interviewing, Testing, and Observation), plus miscellaneous time for each activity to allow for coding, consistency checking, and other similar duties related to the specific activities. Additional provisions are made for time to attend the two training conferences, as well as time allocated to general administrative and coordination activities.

Using these estimates as a guide it appears reasonable, to assume that one person, almost full time, during the 1976-77 school year will suffice for each site. Modifications in this estimate must, of course, be made at the time the precise sample of programs and sites is finally selected. In light of the general oversupply of teachers in the United States, it is highly likely that qualified local data collectors can be identified and enlisted for approximately a regular teacher's contract period and salary.

A number of channels exist for the recruitment of qualified field staff. The primary source, of course, would be the personnel office of the local school district. In general, these personnel offices have far more applications than available openings. Districts may also be willing to give a year's leave of absence to a particularly qualified staff member wishing to participate in the study. Another possible source of field staff would be the education departments of local colleges and universities, which could identify current graduates and other local people active in the educational community. A third source of information on potential field staff candidates would be professional contacts of members of the implementation contractor's project team.

Procedurally, candidates for field staff positions should be identified as soon as possible -- even before final site selection is made. Resumes of potential candidates should be solicited from all sources as the final sample selection process is continuing. The verification site visits to be made prior to final selection represent a good opportunity for the implementation contractor to interview candidates. Immediately upon completion of the selection of sites, (by July 1976) commitments should be made and contracts negotiated.

Training of Field Staff

In order for the study to have an effective data collection arm, the training of field data collectors is crucial. A great deal of time will be spent by the project team (the Training Director, in particular) in developing a high-quality presentation of two training devices, agendas for the two training sessions and a Data Collector's Manual. These two training devices will be parallel in nature, the first serving as initial lesson and the second as a continuing guide throughout the project.

Each of these two training tools will be divided into four basic subsections:

- General information -- in which the purpose of the study is outlined and the role of the local data collector is specified,
- Protocol -- in which political considerations are described and appropriate procedures detailed,
- Administrative -- in which the basic reporting, record-keeping, confidentiality, and communication aspects of the project are described, with particular emphasis on the written requirements and administrative logs (see Task 6),
- Data collection -- in which each of the specific data collection instruments is described and the procedures for its use detailed.

Appendix 14 of this report is a draft of a suggested Data Collector's Manual. In Appendix 14, some specific procedures are described and, where the particular instruments are not completely defined, general procedures are outlined. In addition to the information contained in the Data Collector's Manual, the training sessions will contain provision for a great deal of practice in using each type of instrument. This practice, under simulated classroom and interview situations, is essential if inter-rater reliability is to be high.

In order to meet the tight time constraints imposed by the project schedule and to ensure the highest possible degree of data reliability, the training of local data collectors will be conducted in two five-day training sessions during August 1976. The first session will accentuate general, protocol, and administrative aspects, as well as interview and testing instruments and procedures; the second will concentrate heavily on classroom observation procedures. The locations for the two training sessions will be determined once the specific study sites are selected. They will be chosen for their central location or for other economic factors.

The training received by local data collectors will be reinforced by the Data Collector's Manual and by close interaction between the local data collectors and their designated Regional Coordinator. As indicated in our discussion of Task 6, the Regional Coordinators will assist the local data collectors by teaming with them on some of their initial interviews and observations.

TASK 9 -- DEVELOP IMPLEMENTATION SCHEDULE

In this task, we identify the specific tasks associated with the implementation phase of the individualized instruction study and describe the proposed schedule by which the study may be completed before the Congressional deadline of July 1977.

Task Descriptions

Task 1: Select Final Sample -- This task encompasses those subtasks which relate to the selection of the districts, schools, and classrooms to be included in the study.

1.1 Administer Sampling Questionnaires -- Assuming that the sampling questionnaires for district and school-level personnel (see Appendix 5) will be submitted to OMB for clearance prior to implementation contract award, they will be sent to the appropriate LEA personnel in accordance with the procedures described in Task 2.

1.2 Conduct Verification Visits -- For those sites at which additional or verified information is required. On-site visits will be made to complete and verify the data requirements for all candidate programs.

1.3 Select Sample -- Based upon all available information, including existing program documentation, District Survey I data, sampling questionnaire responses, and site visit data, the final sample of study participants will be selected as indicated in the Task 2 discussion.

Task 2: Select and Train Data Collectors -- This task includes all activities associated with the recruitment, selection, and training of the field staff which will perform the on-site data collection duties.

2.1 Interview Local Candidates -- During verification visits (Subtask 1.3), potential candidates for position as local data collector will be solicited and interviewed.

2.2 Select Data Collectors -- When the final selection of sites has occurred (Subtask 1.4), the local data collectors for each selected site will be chosen.

2.3 Develop Training Materials -- The materials and techniques to be used in the training of local data collectors will be developed.

2.4 Conduct Training Session I -- The first training session, including a general overview of the project, interview instruments and techniques, and testing procedures, will be conducted.

2.5 Conduct Training Session II -- The second training session, on classroom observation instruments and procedures will be conducted.

Task 3: Conduct Data Collection -- This task reflects the three basic data collection activities of the study and the corresponding preliminary activities.

3.1 Finalize Data Collection Instruments -- The specific instruments that will be used to collect data (including interview, testing, and classroom observation instruments) will be finalized and, along with their corresponding documentation packages, submitted to OMB for clearance. This subtask also includes the field test of the instruments.

3.2 Prepare for Data Collection -- Contacts and schedules for the conduct of data collection activities are made.

3.3 Conduct Interviews -- Interviews with LEA will be conducted at the beginning of the school year and again toward the end of the school year.

3.4 Conduct Testing -- During the early autumn of 1976, student pre-tests in both achievement and affective areas will be administered. Corresponding post-tests will be administered during April 1977. In both pre and post-tests, particular care will be exercised to ensure effective coordination of such administrative matters as ensuring the appropriate space and developing a quick response coding plan.

3.5 Conduct Observations -- Classroom observations will be conducted throughout the period between September 1976 and April 1977.

Task 4: Conduct Data Analysis -- This task includes the reduction, preparation, and analysis of data. It should be noted that, in order to facilitate the timely analysis of data, data reduction, verification, and analysis will be conducted on a continuing basis during the 1976-77 school year.

4.1 Reduce Data -- The raw data will be converted into a format suitable for analysis.

4.2 Obtain Scored Test Results -- Pre- and post-test scores will be obtained from the scorers.

4.3 Check Data for Consistency -- The reduced data must be checked for internal consistency prior to analysis.

4.4 Conduct Specific Analyses -- The analysis activities outlined in the analysis plan (Section III/Task 5), will be conducted.

Task 5: Prepare Final Report -- This task reflects the preparation of a final report to NIE.

5.1 Prepare Draft Final Report -- The final report will be prepared in draft form.

5.2 Prepare Final Report -- Based upon NIE recommendations, the draft final report will be revised into the final version.

Project Schedule

Exhibit IV-5 displays, in the form of a GANTT chart, the schedule by which the project tasks will be conducted. Exhibit IV-6 expands the schedule into the form of a PERT network. The final chart, Exhibit IV-7, describes the specific products and reports to be submitted during the implementation phase of the study.

EXHIBIT IV-5

GANTT CHART FOR PROPOSED STUDY

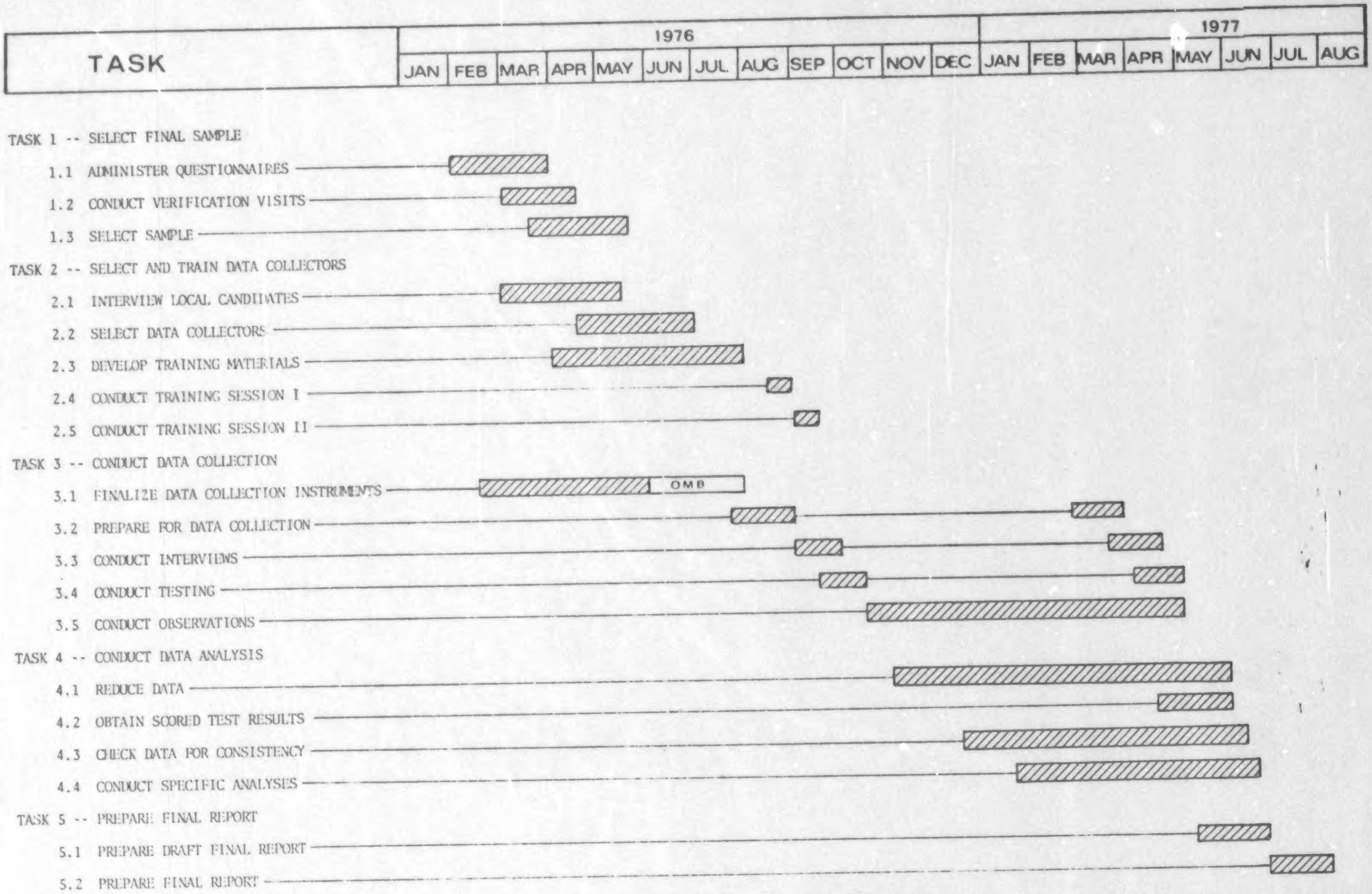
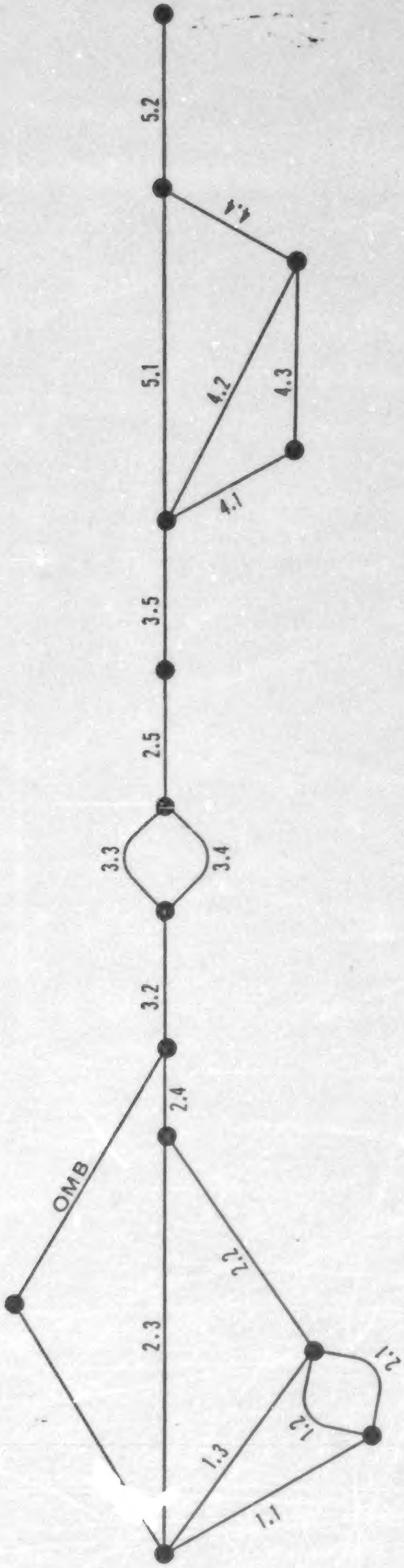


Exhibit IV-6

PERT CHART FOR PROPOSED STUDY


5.2
4.4
5.1
4.2
4.3
4.1
3.5
2.5
3.3
3.4
3.2
2.4
OMB
2.2
2.3
2.1
1.2
1.3
1.1

Numbers refer to tasks as shown on Exhibit IV-5

EXHIBIT IV-7

LIST OF DELIVERABLE PRODUCTS

	Item	Date
1.	Final Data Collection Instruments	1 June 1976
	• Interview Instruments • Observation Instruments • Achievement Tests • Affective Instruments • Complete OMB Documentation	
2.	Training Materials	1 August 1976
	• Training Conference Agendae • Data Collector's Manual	
3.	Interim Report	10 November 1976
	• Project Status Report • Interview Summary • Testing Summary	
3.	Draft Final Report	20 June 1977
	• Project Summary • Analysis Findings • Recommendations	
5.	Final Report	31 July 1977

Monthly Progress Reports, in letter form, will be submitted throughout the project period.